(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

AC

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 January 2001 (11.01.2001)

PCT

(10) International Publication Number
WO 01/02568 A2

(51) International Patent Classification[7]: C12N 15/12, 15/55, 15/54, 15/61, C07K 14/47, C12N 9/64, 9/12, 9/90, C12Q 1/68, C12N 15/11, C07K 16/18, 16/40, G01N 33/566, A61K 38/00

(21) International Application Number: PCT/US00/18374

(22) International Filing Date: 30 June 2000 (30.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/142,310	2 July 1999 (02.07.1999)	US
60/142,311	2 July 1999 (02.07.1999)	US

(71) Applicants: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US). **HYSEQ, INC.** [US/US]; 675 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors: **WILLIAMS, Lewis, T.**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **ESCOBEDO, Jaime**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **INNIS, Michael, A.**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **GARCIA, Pablo, Dominguez**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **KLINGER, Julie**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **KASSAM, Altaf**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **REINHARD, Christoph**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **RANDAZZO, Filippo**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **KENNEDY, Guilia, C.**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **POT, David**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **LAMSON, George**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **DRMANAC, Radoje**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **CRKENJAKOV, Radomir**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **DRMANAC, Snezana**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **DICKSON, Mark**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **LABAT, Ivan**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **LESHKOWITIZ, Dena**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **KITA, David**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **GARCIA, Veronica**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **JONES, Lee, William**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US). **STRACHE-CRAIN, Birgit**; 675 Almanor Avenue, Sunnyvale, CA 94086 (US).

(74) Agents: **BLACKBURN, Robert, P.**; Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608-2916 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/02568 A2

(54) Title: NOVEL HUMAN GENES AND GENE EXPRESSION PRODUCTS

(57) Abstract: The invention provides novel polynucleotides. The invention further provides novel members of protein families, and polynucleotides that are differentially expressed in cancer cells relative to normal cells, and in metastatic cancer cells relative to normal cells or non-metastatic cancer cells.

NOVEL HUMAN GENES AND GENE EXPRESSION PRODUCTS

FIELD OF THE INVENTION

The present invention relates to novel polynucleotides of human origin and the encoded gene products.

BACKGROUND OF THE INVENTION

Identification of novel polynucleotides, particularly those that encode an expressed gene product, is important in the advancement of drug discovery, diagnostic technologies, and the understanding of the progression and nature of complex diseases such as cancer. Identification of genes expressed in different cell types isolated from sources that differ in disease state or stage, developmental stage, exposure to various environmental factors, the tissue of origin, the species from which the tissue was isolated, and the like is key to identifying the genetic factors that are responsible for the phenotypes associated with these various differences.

This invention provides novel human polynucleotides, the polypeptides encoded by these polynucleotides, and the genes and proteins corresponding to these novel polynucleotides.

SUMMARY OF THE INVENTION

This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides and to proteins expressed by the genes. The invention also relates to diagnostics and therapeutics comprising such novel human polynucleotides, their corresponding genes or gene products, including probes, antisense nucleotides, and antibodies. The polynucleotides of the invention correspond to a polynucleotide comprising the sequence information of at least one of SEQ ID NOs: 1-3351.

Various aspects and embodiments of the invention will be readily apparent to the ordinarily skilled artisan upon reading the description provided herein.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to polynucleotides comprising the disclosed nucleotide sequences, to full length cDNA, mRNA genomic sequences, and genes

corresponding to these sequences and degenerate variants thereof, and to polypeptides encoded by the polynucleotides of the invention and polypeptide variants.

Polypeptide variants differ from wild type protein in having one or more amino acid substitutions that either enhance, add, or diminish a biological activity of the wild type protein.

Six of the polypeptides disclosed herein encode new members of the MKK kinase family; the coding region is found within the nucleotide region in parentheses: SEQ ID NO:29 (nucleotides 295-421); SEQ ID NO:31 (298-397); SEQ ID NO:196 (37-322); SEQ ID NO:3175 (nucleotides 14-164); SEQ ID NO:3190 (229-390); and SEQ ID NO:3281 (15-182). Twenty-four of the polypeptides encode new members of the family of transcription factor proteins having a basic region plus leucine zipper: SEQ ID NO:410 (42-191); SEQ ID NO:552 (116-288); SEQ ID NO:768 (116-288); SEQ ID NO:822 (108-262); SEQ ID NO:836 (158-353); SEQ ID NO:1288 (73-234); SEQ ID NO:1365 (69-257); SEQ ID NO:1540 (289-471); SEQ ID NO:1549 (200-391); SEQ ID NO:1556 (163-354); SEQ ID NO:1557 (207-398); SEQ ID NO:1563 (107-298); SEQ ID NO:1622 (180-365); SEQ ID NO:1630 (100-291); SEQ ID NO:1704 (184-372); SEQ ID NO:1808 (36-161); SEQ ID NO:1454 (49-209); SEQ ID NO:2363 (48-211); SEQ ID NO:2424 (43-194); SEQ ID NO:3147 (190-369); SEQ ID NO:3152 (129-320); SEQ ID NO:3158 (167-334); and SEQ ID NO:3208 (34-256).

SEQ ID NOs:186 (175-395); 2591 (60-165); 3307 (43-321); and 3339 (94-342) encode polypeptides having an SH2 domain, and SEQ ID NOs:234 (23-121), 1832 (18-173), and 1835 (57-206) encode polypeptides having an SH3 domain. Nine polypeptides encode new members of the family of proteins having Ank repeat regions: SEQ ID NO:187 (358-432); SEQ ID NO:1268 (238-315); SEQ ID NO:1804 (301-378); SEQ ID NO:1819 (278-355); SEQ ID NO:1839 (224-307); SEQ ID NO:1830 (184-267); SEQ ID NO:2562 (18-101); SEQ ID NO:3015 (131-214); and SEQ ID NO:3267 (97-180).

The following eleven polynucleotides encode polypeptides having a C2H2 type zinc finger: SEQ ID NOs:308 (110-172); 807 (339-392); 1324 (294-356); 1503 (154-216); 1527 (156-212); 1674 (196-258); 1779 (64-126); 1801 (295-351); 3081 (190-252); 3193 (293-355); and 3306 (161-223). Eight polynucleotides encode polypeptides of the family of ATPases: SEQ ID NOs:431 (71-428); 639 (157-561); 2135 (2-401); 2684 (9-461); 2859 (100-320); 3178 (45-386); 3197 (281-343) and 3266 (8-139). Polypeptides having a fibronectin type III domain are encoded by SEQ ID NO:746 (209-427) and 1192 (186-416). Polypeptides having an EF-hand domain are encoded by SEQ ID NO:820 (341-

406); 1755 (281-367) and 3285(16-102). Six polypeptides of the protein kinase family are encoded by SEQ ID NOs:1157 (41-444); 1478 (54-437), 1496 (241-520); 2286 (12-182); 2969 (5-387); and 3190 (118-390).

LIM domain-containing polypeptides are encoded by SEQ ID NO:1269 (79-240); 1309 (248-404); 1360 (222-377); and 1386 (243-398). Two polypeptides of the family having a C2 domain (protein kinase C-like) are encoded by SEQ ID NO:1325 (1-234) and 2282(183-353). Polypeptides having a WD domain, G-beta repeat motif are encoded by SEQ ID NOs:1336 (66-164); 1380 (42-140); 1711 (263-361); 1762 (236-334); 1909 (160-258); 2218 (127-225); 3047 (191-292); 3108 (275-367) and 3292 (208-300).

SEQ ID NO:1410 (222-350) encodes a member of the trypsin family. SEQ ID NOs:1417 (8-354); 2281 (20-387) and 2310 (20-371) encode members of the protein tyrosine phosphatase family. SEQ ID NOs:1464 (4-180) and 1514 (2-252) encode members of the family having an RNA recognition motif (also known as RRM, RBD, or RNP domain). SEQ ID NOs:1496 (241-520) and 3297(7-153) encode helicases having a conserved C-terminal domain. SEQ ID NO:1538 (9-635) encodes a member of the wnt family of developmental signaling proteins.

Three polynucleotides encode polypeptides having a homeobox domain: SEQ ID NOs:1676 (9-86); 1820 (123-299); and 1821 (127-303). A novel thioredoxin is encoded by SEQ ID NO:1677 (316-369). Two novel members of the ras family are encoded by SEQ ID NO:1688(109-410) and 3258(138-394). A novel polypeptide having a phosphatidylinositol-specific phospholipase C Y-domain is encoded by SEQ ID NO:1707 (92-439). A novel serine carboxypeptidase is encoded by SEQ ID NO:1744 (238-433). A novel polypeptide having N-terminal homology in the Ets domain is encoded by SEQ ID NO:1811 (184-315). A novel polypeptide having a bromodomain is encoded by SEQ ID NO:1814 (127-294). A novel polypeptide having a double-stranded RNA binding motif is encoded by SEQ ID NO:1818 (9-146). A novel polypeptide having a G-protein alpha subunit is encoded by SEQ ID NO:1846 (12-398).

SEQ ID NOs:1911 (35-151) and 1980 (60-197) encode polypeptides having a C3HC4 type zinc finger domain (RING finger). SEQ ID NO:2065 (253-306) encodes a polypeptide having a CCHC zinc finger domain. SEQ ID NO:2216 (90-179) encodes a polypeptide having a WW/rsp5/WWP domain. SEQ ID NO:2428 (25-350) encodes a polypeptide member of the dual specificity phosphatase family, having a catalytic domain.

SEQ ID NOs:2577 (0-311); 3183 (14-215); and 3195 (0-215) encode members of the 4 transmembrane segment integral membrane protein family. SEQ ID

NOs:2826 (116-400) and 2871 (198-392) encode polypeptides of the DEAD and DEAH box helicase family. SEQ ID NO:2944 (18-281) encodes a polypeptide having a calpain large subunit, domain III.

SEQ ID NO:3274 (11-187) encodes a eukaryotic transcription factor with a fork head domain. SEQ ID NO:3345 (65-271) encodes a polypeptide having a PDZ domain, and SEQ ID NO:3351 (124-270) encodes a polypeptide in the family of phorbol esters/glycerol binding proteins.

Described below are polynucleotide compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these polynucleotides and genes, identification of structural motifs of the polynucleotides and genes, identification of the function of a gene product encoded by a gene corresponding to a polynucleotide of the invention, use of the provided polynucleotides as probes and in mapping and in tissue profiling, use of the corresponding polypeptides and other gene products to raise antibodies, and use of the polynucleotides and their encoded gene products for therapeutic and diagnostic purposes.

Polynucleotide Compositions

The scope of the invention with respect to polynucleotide compositions includes, but is not necessarily limited to, polynucleotides having a sequence set forth in any one of SEQ ID NOs:1-3351; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here. "Polynucleotide" and "nucleic acid" as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicated.

The invention features polynucleotides that are expressed in human tissue, specifically human colon, breast, and/or lung tissue. Novel nucleic acid

compositions of the invention comprise a sequence set forth in any one of SEQ ID NOs:1-3351 or an identifying sequence thereof. An "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a polynucleotide sequence, *e.g.*, exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of SEQ ID NOs:1-3351.

The polynucleotides of the invention also include polynucleotides having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, U.S. Patent No. 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.*, allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences (SEQ ID NOs:1-3351) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, *e.g.*, primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

Preferably, hybridization is performed using at least 15 contiguous nucleotides (nt) of at least one of SEQ ID NOs:1-3351. That is, when at least 15 contiguous nt of one of the disclosed SEQ ID NOs. is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one SEQ ID NO. can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA. Probes of more than 15 nt can be used, *e.g.*, probes of from about 18 nt to about 100 nt, but 15 nt represents sufficient sequence for unique identification.

The polynucleotides of the invention also include naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants, allelic variants).

Variants of the polynucleotides of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

The invention also encompasses homologs corresponding to the polynucleotides of SEQ ID NOs:1-3351, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.* Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al., *J. Mol. Biol.* (1990) *215*:403-10.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90%, 91%, 92%, 93%, 94%, 95%, or 96%, most preferably 97%, 98% or 99%. For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (*e.g.*, in diagnosis, as a unique identifier of a differentially expressed gene of interest, *etc.*). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence

elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

The nucleic acid compositions of the subject invention can encode all or a part of the subject polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* Isolated polynucleotides and polynucleotide fragments of the invention comprise at least about 10, about 15, about 20, about 35, about 50, about 100, about 150 to about 200, about 250 to about 300, or about 350 contiguous nt selected from the polynucleotide sequences as shown in SEQ ID NOs:1-3351. The fragments also include those of lengths intermediate to the specifically mentioned lengths, such as 35, 36, 37, 38, 39, etc.; 150, 151, 152, 153, 154, etc. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more. In a preferred embodiment, the polynucleotide molecules comprise a contiguous sequence of at least 12 nt selected from the group consisting of the polynucleotides shown in SEQ ID NOs:1-3351.

Probes specific to the polynucleotides of the invention can be generated using the polynucleotide sequences disclosed in SEQ ID NOs:1-3351. The probes are preferably at least about a 12, 15, 16, 18, 20, 22, 24, or 25 nt fragment of a corresponding contiguous sequence of SEQ ID NOs:1-3351, and can be less than 2, 1, 0.5, 0.1, or 0.05 kb in length. The probes can be synthesized chemically or can be generated from longer polynucleotides using restriction enzymes. The probes can be

labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a polynucleotide of one of SEQ ID NOs:1-3351. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

The polynucleotides of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The polynucleotides of the invention can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides of the invention can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological samples (*e.g.*, extracts of human cells) to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of the polynucleotide sequences as shown in SEQ ID NOs:1-3351 or variants thereof in a sample. These and other uses are described in more detail below.

Use of Polynucleotides to Obtain Full-Length cDNA, Gene, and Promoter Region

Full-length cDNA molecules comprising the disclosed polynucleotides are obtained as follows. A polynucleotide having a sequence of one of SEQ ID NOs:1-3351, or a portion thereof comprising at least 12, 15, 18, or 20 nt, is used as a hybridization probe to detect hybridizing members of a cDNA library using probe design methods, cloning methods, and clone selection techniques such as those described in U.S. Patent No. 5,654,173. Libraries of cDNA are made from selected tissues, such as normal or tumor tissue, or from tissues of a mammal treated with, for example, a pharmaceutical agent. Preferably, the tissue is the same as the tissue from which the polynucleotides of the invention were isolated, as both the polynucleotides described herein and the cDNA represent expressed genes. Most preferably, the cDNA library is made from the biological material described herein in the Examples. The choice of cell type for library construction can be made after the identity of the protein encoded by the gene corresponding to the polynucleotide of the invention is known. This will indicate which tissue and cell types are likely to express the related gene, and thus represent a suitable source for the mRNA for generating the cDNA. As described in the Examples, cDNA of the invention was isolated from specific cell or tissue types, and such cells and tissues are preferable for obtaining related nucleic acids.

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. The cDNA can be prepared by using primers based on sequence from SEQ ID NOs:1-3351. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided polynucleotides, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. In order to obtain additional sequences

5' to the end of a partial cDNA, 5' RACE (*PCR Protocols: A Guide to Methods and Applications*, (1990) Academic Press, Inc.) can be performed.

Genomic DNA is isolated using the provided polynucleotides in a manner similar to the isolation of full-length cDNAs. Briefly, the provided polynucleotides, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the polynucleotides of the invention, but this is not essential. Most preferably, the genomic DNA is obtained from the biological material described herein in the Examples. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook et al., 9.4-9.30. In addition, genomic sequences can be isolated from human BAC libraries, which are commercially available from Research Genetics, Inc., Huntsville, Alabama, USA, for example. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook et al., such that adjacent and overlapping fragments of genomic DNA are isolated. These are mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

Using the polynucleotide sequences of the invention, corresponding full-length genes can be isolated using both classical and PCR methods to construct and probe cDNA libraries. Using either method, Northern blots, preferably, are performed on a number of cell types to determine which cell lines express the gene of interest at the highest level. Classical methods of constructing cDNA libraries are taught in Sambrook et al., *supra*. With these methods, cDNA can be produced from mRNA and inserted into viral or expression vectors. Typically, libraries of mRNA comprising poly(A) tails can be produced with poly(T) primers. Similarly, cDNA libraries can be produced using the instant sequences as primers.

PCR methods are used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant polynucleotides. Such PCR methods include gene trapping and RACE methods as described in Gruber et al., WO 95/04745 and Gruber et al., U.S. Patent No. 5,500,356. Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA. In preferred embodiments of RACE, a common primer is designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends (Apte and Siebert, *Biotechniques* (1993) *15*:890-893; Edwards et al., *Nuc. Acids Res.* (1991) *19*:5227-5232). When a single gene-specific RACE primer is paired with the common

primer, preferential amplification of sequences between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

The promoter region of a gene generally is located 5' to the initiation site for RNA polymerase II. Hundreds of promoter regions contain the "TATA" box, a sequence such as TATTA or TATAA, which is sensitive to mutations. The promoter region can be obtained by performing 5' RACE using a primer from the coding region of the gene. Alternatively, the cDNA can be used as a probe for the genomic sequence, and the region 5' to the coding region is identified by "walking up." If the gene is highly expressed or differentially expressed, the promoter from the gene can be of use in a regulatory construct for a heterologous gene.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook et al., 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure herein on optional changes in amino acids to achieve altered protein structure and/or function.

As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more polynucleotides of the invention can be synthesized. Thus, the invention encompasses nucleic acid molecules ranging in length from 15 nt (corresponding to at least 15 contiguous nt of one of SEQ ID NOs:1-3351) up to a maximum length suitable for one or more biological manipulations, including replication and expression, of the nucleic acid molecule. The invention includes but is not limited to (a) nucleic acid having the size of a full gene, and comprising at least one of SEQ ID NOs:1-3351; (b) the nucleic acid of (a) also comprising at least one additional polynucleotide or gene, operably linked to permit expression of a fusion protein; (c) an expression vector comprising (a) or (b); (d) a plasmid comprising (a) or (b) ; and (e) a recombinant viral particle comprising (a) or (b). Once provided with the polynucleotides disclosed herein, construction or preparation of (a) - (e) are well within the skill in the art.

The sequence of a nucleic acid comprising at least 15 contiguous nt of at least any one of SEQ ID NOs:1-3351, preferably the entire sequence of at least any one of SEQ ID NOs:1-3351, is not limited and can be any sequence of A, T, G, and/or C (for DNA) and A, U, G, and/or C (for RNA) or modified bases thereof, including inosine and pseudouridine. The choice of sequence will depend on the desired function and can be dictated by coding regions desired, the intron-like regions desired, and the

regulatory regions desired. Where the entire sequence of any one of SEQ ID NOs:1-3351 is within the nucleic acid, the nucleic acid obtained is referred to herein as a polynucleotide comprising the sequence of any one of SEQ ID NOs:1-3351.

Expression of Polypeptide Encoded by Full-Length cDNA or Full-Length Gene

The provided polynucleotides (*e.g.*, a polynucleotide having a sequence of one of SEQ ID NOs:1-3351), the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having sequences of SEQ ID NOs:1-3351 can be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.*, Stemmer et al., *Gene* (*Amsterdam*) (1995) *164*(*1*):49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, *Nature* (1994) *370*:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process.

Appropriate polynucleotide constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. The gene product encoded by a polynucleotide of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Vectors, host cells and methods for obtaining expression in same are well known in the art. Suitable vectors and host cells are described in U.S. Patent No. 5,654,173.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are

available commercially. Methods for preparation of vectors comprising a desired sequence are well known in the art.

The polynucleotides set forth in SEQ ID NOs:1-3351 or their corresponding full-length polynucleotides are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

When any appropriate host cells or organisms are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence as disclosed in U.S. Patent No. 5,641,670.

Identification of Functional and Structural Motifs of Novel Genes

Translations of the nucleotide sequence of the provided polynucleotides, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the polynucleotides of the invention. Also, sequences exhibiting similarity with more than one individual sequence can exhibit activities that are characteristic of either or both individual sequences.

The full length sequences and fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence corresponding to provided polynucleotides. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences corresponding to the provided polynucleotides.

Typically, a selected polynucleotide is translated in all six frames to determine the best alignment with the individual sequences. The sequences disclosed

herein in the Sequence Listing are in a 5' to 3' orientation and translation in three frames can be sufficient. These amino acid sequences are referred to, generally, as query sequences, which will be aligned with the individual sequences. Databases with individual sequences are described in "Computer Methods for Macromolecular Sequence Analysis" *Methods in Enzymology* (1996) *266*, Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Databases include Genbank, EMBL, and DNA Database of Japan (DDBJ).

Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST, available over the world wide web at http://www.ncbi.nlm.nhi.gov/BLAST. Another alignment algorithm is Fasta, available in the Genetics Computing Group (GCG) package, Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Doolittle, *supra*. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Meth. Mol. Biol.* (1997) *70*: 173-187. Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to identify sequences that are distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Amino acid sequences encoded by the provided polynucleotides can be used to search both protein and DNA databases.

High Similarity. In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. If high similarity is found, the query sequence is considered to have high similarity with a profile

sequence when the p value is less than or equal to about 10^{-2}; more usually; less than or equal to about 10^{-3}; even more usually; less than or equal to about 10^{-4}. More typically, the p value is no more than about 10^{-5}; more typically; no more than or equal to about 10^{-10}; even more typically; no more than or equal to about 10^{-15} for the query sequence to be considered high similarity.

Similarity Determined by Sequence Identity Alone. Sequence identity alone can be used to determine similarity of a query sequence to an individual sequence and can indicate the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%. Sequence identity alone as a measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, 90 residues; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in length.

Alignments with Profile and Multiple Aligned Sequences. Translations of the provided polynucleotides can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided polynucleotides can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene products (*e.g.*, polypeptides) encoded by the provided polynucleotides or corresponding cDNA or genes. For example, sequences that show an identity or similarity with a chemokine profile or MSA can exhibit chemokine activities.

Profiles can be designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney et al., *Nucl. Acid Res.* (1996) *24(14)*: 2730-2739. MSAs of some protein families and motifs are publicly available. MSAs are described also in Sonnhammer et al., *Proteins* (1997) *28*: 405-420. A brief description of MSAs is reported in Pascarella et al., *Prot. Eng.* (1996) *9(3)*:249-251. Techniques for building profiles from MSAs are described in Sonnhammer et al., *supra*; Birney et al., *supra*;

and "Computer Methods for Macromolecular Sequence Analysis," *Methods in Enzymology* (1996) *266*, Doolittle, Academic Press, Inc., San Diego, California, USA.

Similarity between a query sequence and a protein family or motif can be determined by (a) comparing the query sequence against the profile and/or (b) aligning the query sequence with the members of the family or motif. Typically, a program such as Searchwise is used to compare the query sequence to the statistical representation of the multiple alignment, also known as a profile (see Birney et al., *supra*). Other techniques to compare the sequence and profile are described in Sonnhammer et al., *supra* and Doolittle, *supra*.

Next, methods described by Feng et al., *J. Mol. Evol.* (1987) 25:351 and Higgins et al., *CABIOS* (1989) *5*:151 can be used align the query sequence with the members of a family or motif, also known as a MSA. Sequence alignments can be generated using any of a variety of software tools. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng et al., *J. Mol. Evol.* (1987) *25*:351. Another method, GAP, uses the alignment method of Needleman et al., *J. Mol. Biol.* (1970) *48*:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith et al., *Adv. Appl. Math.* (1981) *2*:482. In general, the following factors are used to determine if a similarity between a query sequence and a profile or MSA exists: (1) number of conserved residues found in the query sequence, (2) percentage of conserved residues found in the query sequence, (3) number of frameshifts, and (4) spacing between conserved residues.

Some alignment programs that both translate and align sequences can make any number of frameshifts when translating the nucleotide sequence to produce the best alignment. The fewer frameshifts needed to produce an alignment, the stronger the similarity or identity between the query and profile or MSAs. For example, a weak similarity resulting from no frameshifts can be a better indication of activity or structure of a query sequence, than a strong similarity resulting from two frameshifts. Preferably, three or fewer frameshifts are found in an alignment; more preferably two or fewer frameshifts; even more preferably, one or fewer frameshifts; even more preferably, no frameshifts are found in an alignment of query and profile or MSAs.

Conserved residues are those amino acids found at a particular position in all or some of the family or motif members. Alternatively, a position is considered conserved if only a certain class of amino acids is found in a particular position in all or

some of the family members. For example, the N-terminal position can contain a positively charged amino acid, such as lysine, arginine, or histidine.

Typically, a residue of a polypeptide is conserved when a class of amino acids or a single amino acid is found at a particular position in at least about 40% of all class members; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A residue is considered conserved when three unrelated amino acids are found at a particular position in the some or all of the members; more usually, two unrelated amino acids. These residues are conserved when the unrelated amino acids are found at particular positions in at least about 40% of all class member; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A query sequence has similarity to a profile or MSA when the query sequence comprises at least about 25% of the conserved residues of the profile or MSA; more usually, at least about 30%; even more usually; at least about 40%. Typically, the query sequence has a stronger similarity to a profile sequence or MSA when the query sequence comprises at least about 45% of the conserved residues of the profile or MSA; more typically, at least about 50%; even more typically; at least about 55%.

Identification of Secreted and Membrane-Bound Polypeptides

Both secreted and membrane-bound polypeptides of the present invention are of particular interest. For example, levels of secreted polypeptides can be assayed in body fluids that are convenient, such as blood, plasma, serum, and other body fluids such as urine, prostatic fluid and semen. Membrane-bound polypeptides are useful for constructing vaccine antigens or inducing an immune response. Such antigens would comprise all or part of the extracellular region of the membrane-bound polypeptides. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides.

A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods, *Proc. Natl. Acad. Sci. USA* (1981) *78*:3824-3828; Kyte & Doolittle, *J. Mol. Biol.* (1982) *157*: 105-132; and RAOAR algorithm, Degli Esposti et al., *Eur. J. Biochem.* (1990) *190*: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the polynucleotides of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine

Identification of the Function of an Expression Product of a Full-Length Gene

Ribozymes, antisense constructs, and dominant negative mutants can be used to determine function of the expression product of a gene corresponding to a polynucleotide provided herein. The phosphoramidite method of oligonucleotide synthesis can be used to construct antisense molecules and ribozymes. See Beaucage et al., *Tet. Lett.* (1981) *22*:1859 and U.S. Patent No. 4,668,777. Automated devices for synthesis are available to create oligonucleotides using this chemistry. Examples of such devices include Biosearch 8600, Models 392 and 394 by Applied Biosystems, a division of Perkin-Elmer Corp., Foster City, California, USA; and Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Synthetic RNA, phosphate analog oligonucleotides, and chemically derivatized oligonucleotides can also be produced, and can be covalently attached to other molecules. RNA oligonucleotides can be synthesized, for example, using RNA phosphoramidites. This method can be performed on an automated synthesizer, such as Applied Biosystems, Models 392 and 394, Foster City, California, USA.

Oligonucleotides of up to 200 nt can be synthesized, more typically, 100 nt, more typically 50 nt; even more typically 30 to 40 nt. These synthetic fragments can be annealed and ligated together to construct larger fragments. See, for example, Sambrook et al., *supra*. Trans-cleaving catalytic RNAs (ribozymes) are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message must contain a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit expression of a gene of unknown function for the purpose of determining its function in an *in vitro* or *in vivo* context, by detecting the phenotypic effect.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense polynucleotides based on a selected polynucleotide sequence can interfere with expression of the corresponding gene. Antisense polynucleotides are typically generated within the cell by expression from antisense constructs that contain the antisense strand as the transcribed strand. Antisense polynucleotides based on the disclosed polynucleotides will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense polynucleotide. The expression products of control cells and cells treated with the antisense construct are compared to detect the protein product of the gene corresponding to the polynucleotide upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

Given the extensive background literature and clinical experience in antisense therapy, one skilled in the art can use selected polynucleotides of the invention as additional potential therapeutics. The choice of polynucleotide can be narrowed by first testing them for binding to "hot spot" regions of the genome of cancerous cells. If a polynucleotide is identified as binding to a "hot spot," testing the polynucleotide as an antisense compound in the corresponding cancer cells is warranted.

Dominant negative mutations also are readily generated for corresponding proteins that are active as homomultimers. A mutant polypeptide will interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic

domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants (see, *e.g.*, Herskowitz, *Nature* (1987) *329*:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

Polypeptides and Variants Thereof

The polypeptides of the invention include those encoded by the disclosed polynucleotides, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of SEQ ID NOs:1-3351 or a variant thereof.

In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein (*e.g.*, human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide of the invention, as measured by BLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, *i.e.*, the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, *i.e.*, other animal or plant species, where such homologs, usually mammalian species, *e.g.*, rodents, such as mice, rats; domestic animals, *e.g.*, horse, cow, dog, cat; and humans. By "homolog" is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the BLAST algorithm, with the parameters described above.

In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, *e.g.*, are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/ hydrophilicity, and/or steric bulk of the amino acid substituted. Variants can be designed so as to retain biological activity of a particular region of the protein (*e.g.*, a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (see, *e.g.*, Go et al., *Int. J. Peptide Protein Res.* (1980) *15*:211), the thermostability of the variant polypeptide (see, *e.g.*, Querol et al., *Prot. Eng.* (1996) *9*:265), desired glycosylation sites (see, *e.g.*, Olsen and Thomsen, J. *Gen. Microbiol.* (1991) *137*:579), desired disulfide bridges (see, *e.g.*, Clarke et al., *Biochemistry* (1993) *32*:4322; and Wakarchuk et al., *Protein Eng.* (1994) *7*:1379), desired metal binding sites (see, *e.g.*, Toma et al., *Biochemistry* (1991) *30*:97, and Haezerbrouck et al., *Protein Eng.* (1993) *6*:643), and desired substitutions with in proline loops (see, *e.g.*, Masul et al., *Appl. Env. Microbiol.* (1994) *60*:3579). Cysteine-depleted muteins can be produced as disclosed in U.S. Patent No. 4,959,314.

Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a

polynucleotide having a sequence of any SEQ ID NOs:1-3351, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

Computer-Related Embodiments

In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (*e.g.*, as a collection of polynucleotide molecules), or in electronic form (*e.g.*, as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, *e.g.*, as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.*, cell type markers), and/or as markers of a given disease or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (*e.g.*, a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a breast ductal cell affected by cancer relative to a normal (*i.e.*, substantially disease-free) breast cell.

The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.*, electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.*, overexpressed or underexpressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have

the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of SEQ ID NOs:1-3351. By plurality is meant at least 2, usually at least 3 and can include up to all of SEQ ID NOs:1-3351. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.*, if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, *etc.*

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.*, the nucleic acid sequences of any of the polynucleotides of SEQ ID NOs:1-3351, can be recorded on computer readable media, *e.g.*, any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.*, word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.*, searchable files, executable files, *etc.*, including, but not limited to, for example, search program software, *etc.*).

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the BLAST (Altschul et al.,

supra.) and BLAZE (Brutlag et al. *Comp. Chem.* (1993) *17*:203) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.*, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt. A variety of comparing means can be used to accomplish comparison of sequence information from a sample (*e.g.*, to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but arc not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are

not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

As discussed above, the "library" of the invention also encompasses biochemical libraries of the polynucleotides of SEQ ID NOs:1-3351, *e.g.*, collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.*, a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.*, an array) and the like. Of particular interest are nucleic acid arrays in which one or more of SEQ ID NOs:1-3351 is represented on the array. By array is meant an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by SEQ ID NOs:1-3351.

Use of Polynucleotide Probes in Mapping, and in Tissue Profiling

Polynucleotide probes, generally comprising at least 12 contiguous nt of a polynucleotide as shown in the Sequence Listing, are used for a variety of purposes, such as chromosome mapping of the polynucleotide and detection of transcription levels. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences.

Detection of Expression Levels. Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots,

mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization is quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for *in situ* hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluors, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and U.S. Patent No. 5,124,246.

Alternatively, the Polymerase Chain Reaction (PCR) is another means for detecting small amounts of target nucleic acids (see, *e.g.*, Mullis et al., *Meth. Enzymol.* (1987) *155*:335; U.S. Patent No. 4,683,195; and U.S. Patent No. 4,683,202). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers can be composed of sequence within or 3' and 5' to the polynucleotides of the Sequence Listing. Alternatively, if the primers are 3' and 5' to these polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art, *e.g.*, Southern blot. mRNA or cDNA can also be detected by traditional blotting techniques (*e.g.*, Southern blot, Northern blot, *etc.*) described in Sambrook et al., "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (*e.g.*, without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

Mapping. Polynucleotides of the present invention can be used to identify a chromosome on which the corresponding gene resides. Such mapping can be useful in identifying the function of the polynucleotide-related gene by its proximity to other genes with known function. Function can also be assigned to the polynucleotide-related gene when particular syndromes or diseases map to the same chromosome. For example, use of polynucleotide probes in identification and quantification of nucleic acid sequence aberrations is described in U.S. Patent No. 5,783,387. An exemplary mapping method is fluorescence *in situ* hybridization (FISH), which facilitates comparative genomic hybridization to allow total genome assessment of changes in relative copy number of DNA sequences (see, *e.g.*, Valdes et al., *Methods in Molecular*

Biology (1997) *68*:1). Polynucleotides can also be mapped to particular chromosomes using, for example, radiation hybrids or chromosome-specific hybrid panels. See Leach et al., *Advances in Genetics*, (1995) *33*:63-99; Walter et al., *Nature Genetics* (1994) *7*:22; Walter and Goodfellow, *Trends in Genetics* (1992) *9*:352. Panels for radiation hybrid mapping are available from Research Genetics, Inc., Huntsville, Alabama, USA. The statistical program RHMAP can be used to construct a map based on the data from radiation hybridization with a measure of the relative likelihood of one order versus another. RHMAP is available via the world wide web at http://www.sph.umich.edu-/group/statgen/software. In addition, commercial programs are available for identifying regions of chromosomes commonly associated with disease, such as cancer.

Tissue Typing or Profiling. Expression of specific mRNA corresponding to the provided polynucleotides can vary in different cell types and can be tissue-specific. This variation of mRNA levels in different cell types can be exploited with nucleic acid probe assays to determine tissue types. For example, PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes substantially identical or complementary to polynucleotides listed in the Sequence Listing can determine the presence or absence of the corresponding cDNA or mRNA.

Tissue typing can be used to identify the developmental organ or tissue source of a metastatic lesion by identifying the expression of a particular marker of that organ or tissue. If a polynucleotide is expressed only in a specific tissue type, and a metastatic lesion is found to express that polynucleotide, then the developmental source of the lesion has been identified. Expression of a particular polynucleotide can be assayed by detection of either the corresponding mRNA or the protein product.

Use of Polymorphisms. A polynucleotide of the invention can be used in forensics, genetic analysis, mapping, and diagnostic applications where the corresponding region of a gene is polymorphic in the human population. Any means for detecting a polymorphism in a gene can be used, including, but not limited to electrophoresis of protein polymorphic variants, differential sensitivity to restriction enzyme cleavage, and hybridization to allele-specific probes.

Antibody Production

Expression products of a polynucleotide of the invention, as well as the corresponding mRNA, cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional

method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an *in vitro* expression system.

Methods for production of monoclonal and polyclonal antibodies that specifically bind a selected antigen are well known in the art. The antibodies specifically bind to epitopes present in the polypeptides encoded by polynucleotides disclosed in the Sequence Listing. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. Epitopes that involve non-contiguous amino acids may require a longer polypeptide, *e.g.*, at least 15, 25, or 50 amino acids. Antibodies that specifically bind to human polypeptides encoded by the provided polynucleotides should provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in Western blots or other immunochemical assays. Preferably, antibodies that specifically polypeptides of the invention do not bind to other proteins in immunochemical assays at detectable levels and can immunoprecipitate the specific polypeptide from solution.

The invention also contemplates naturally occurring antibodies specific for a polypeptide of the invention. For example, serum antibodies to a polypeptide of the invention in a human population can be purified by methods well known in the art, *e.g.*, by passing antiserum over a column to which the corresponding selected polypeptide or fusion protein is bound. The bound antibodies can then be eluted from the column, for example using a buffer with a high salt concentration.

In addition to the antibodies discussed above, the invention also contemplates genetically engineered antibodies, antibody derivatives (*e.g.*, single chain antibodies, antibody fragments (*e.g.*, Fab, *etc.*)), according to methods well known in the art.

Other embodiments of the present invention include humanized monoclonal antibodies capable of binding to the polypeptides of the invention. The phrase "humanized antibody" refers to an antibody derived from a non-human antibody - typically a mouse monoclonal antibody. Alternatively, a humanized antibody may be derived from a chimeric antibody that retains or substantially retains the antigen-binding properties of the parental, non-human, antibody but which exhibits diminished immunogenicity as compared to the parental antibody when administered to humans. The phrase "chimeric antibody," as used herein, refers to an antibody containing

sequence derived from two different antibodies (*see, e.g.,* U.S. Patent No. 4,816,567) which typically originate from different species. Most typically, chimeric antibodies comprise human and murine antibody fragments, generally human constant and mouse variable regions.

Because humanized antibodies are far less immunogenic in humans than the parental mouse monoclonal antibodies, they can be used for the treatment of humans with far less risk of anaphylaxis. Thus, these antibodies may be preferred in therapeutic applications that involve *in vivo* administration to a human such as, *e.g.*, use as radiation sensitizers for the treatment of neoplastic disease or use in methods to reduce the side effects of, *e.g.*, cancer therapy.

Humanized antibodies may be achieved by a variety of methods including, for example: (1) grafting the non-human complementarity determining regions (CDRs) onto a human framework and constant region (a process referred to in the art as "humanizing"), or, alternatively, (2) transplanting the entire non-human variable domains, but "cloaking" them with a human-like surface by replacement of surface residues (a process referred to in the art as "veneering"). In the present invention, humanized antibodies will include both "humanized" and "veneered" antibodies. These methods are disclosed in, *e.g.*, Jones et al., *Nature 321*:522-525 (1986); Morrison et al., *Proc. Natl. Acad. Sci., U.S.A., 81*:6851-6855 (1984); Morrison and Oi, *Adv. Immunol., 44*:65-92 (1988); Verhoeyer et al., *Science 239*:1534-1536 (1988); Padlan, *Molec. Immun. 28*:489-498 (1991); Padlan, *Molec. Immunol. 31(3)*:169-217 (1994); and Kettleborough, C.A. et al., *Protein Eng. 4(7)*:773-83 (1991) each of which is incorporated herein by reference.

The phrase "complementarity determining region" refers to amino acid sequences which together define the binding affinity and specificity of the natural Fv region of a native immunoglobulin binding site. *See, e.g.,* Chothia et al., *J. Mol. Biol. 196*:901-917 (1987); Kabat et al., U.S. Dept. of Health and Human Services NIH Publication No. 91-3242 (1991). The phrase "constant region" refers to the portion of the antibody molecule that confers effector functions. In the present invention, mouse constant regions are substituted by human constant regions. The constant regions of the subject humanized antibodies are derived from human immunoglobulins. The heavy chain constant region can be selected from any of the five isotypes: alpha, delta, epsilon, gamma or mu.

One method of humanizing antibodies comprises aligning the non-human heavy and light chain sequences to human heavy and light chain sequences,

selecting and replacing the non-human framework with a human framework based on such alignment, molecular modeling to predict the conformation of the humanized sequence and comparing to the conformation of the parent antibody. This process is followed by repeated back mutation of residues in the CDR region which disturb the structure of the CDRs until the predicted conformation of the humanized sequence model closely approximates the conformation of the non-human CDRs of the parent non-human antibody. Such humanized antibodies may be further derivatized to facilitate uptake and clearance, *e.g.*, via Ashwell receptors. *See, e.g.,* U.S. Patent Nos. 5,530,101 and 5,585,089 which patents are incorporated herein by reference.

Humanized antibodies can also be produced using transgenic animals that are engineered to contain human immunoglobulin loci. For example, WO 98/24893 discloses transgenic animals having a human Ig locus wherein the animals do not produce functional endogenous immunoglobulins due to the inactivation of endogenous heavy and light chain loci. WO 91/10741 also discloses transgenic non-primate mammalian hosts capable of mounting an immune response to an immunogen, wherein the antibodies have primate constant and/or variable regions, and wherein the endogenous immunoglobulin-encoding loci are substituted or inactivated. WO 96/30498 discloses the use of the Cre/Lox system to modify the immunoglobulin locus in a mammal, such as to replace all or a portion of the constant or variable region to form a modified antibody molecule. WO 94/02602 discloses non-human mammalian hosts having inactivated endogenous Ig loci and functional human Ig loci. U.S. Patent No. 5,939,598 discloses methods of making transgenic mice in which the mice lack endogenous heavy claims, and express an exogenous immunoglobulin locus comprising one or more xenogeneic constant regions.

Using a transgenic animal described above, an immune response can be produced to a selected antigenic molecule, and antibody-producing cells can be removed from the animal and used to produce hybridomas that secrete human monoclonal antibodies. Immunization protocols, adjuvants, and the like are known in the art, and are used in immunization of, for example, a transgenic mouse as described in WO 96/33735. This publication discloses monoclonal antibodies against a variety of antigenic molecules including IL-6, IL-8, TNF , human CD4, L-selectin, gp39, and tetanus toxin. The monoclonal antibodies can be tested for the ability to inhibit or neutralize the biological activity or physiological effect of the corresponding protein. WO 96/33735 discloses that monoclonal antibodies against IL-8, derived from immune cells of transgenic mice immunized with IL-8, blocked IL-8-induced functions of

neutrophils. Human monoclonal antibodies with specificity for the antigen used to immunize transgenic animals are also disclosed in WO 96/34096.

Polynucleotides or Arrays for Diagnostics

Polynucleotide arrays are created by spotting polynucleotide probes onto a substrate (*e.g.*, glass, nitrocellose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (*e.g.*, using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Techniques for constructing arrays and methods of using these arrays are described in EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; U.S. Patent No. 5,593,839; U.S. Patent No. 5,578,832; EP 728 520; U.S. Patent No. 5,599,695; EP 721 016; U.S. Patent No. 5,556,752; WO 95/22058; and U.S. Patent No. 5,631,734. Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a polynucleotide between a test cell and control cell (*e.g.*, cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado et al., *Sem. Radiation Oncol.* (1998) *8*:217; and Ramsay, *Nature Biotechnol.* (1998) *16*:40.

Differential Expression in Diagnosis

The polynucleotides of the invention can also be used to detect differences in expression levels between two cells, *e.g.*, as a method to identify abnormal or diseased tissue in a human. For polynucleotides corresponding to profiles of protein families, the choice of tissue can be selected according to the putative biological function. In general, the expression of a gene corresponding to a specific polynucleotide is compared between a first tissue that is suspected of being diseased and a second, normal tissue of the human. The tissue suspected of being abnormal or diseased can be derived from a different tissue type of the human, but preferably it is derived from the same tissue type; for example an intestinal polyp or other abnormal growth should be compared with normal intestinal tissue. The normal tissue can be the

same tissue as that of the test sample, or any normal tissue of the patient, especially those that express the polynucleotide-related gene of interest (*e.g.*, brain, thymus, testis, heart, prostate, placenta, spleen, small intestine, skeletal muscle, pancreas, and the mucosal lining of the colon). A difference between the polynucleotide-related gene, mRNA, or protein in the two tissues which are compared, for example in molecular weight, amino acid or nucleotide sequence, or relative abundance, indicates a change in the gene, or a gene which regulates it, in the tissue of the human that was suspected of being diseased. Examples of detection of differential expression and its use in diagnosis of cancer are described in U.S. Patent Nos. 5,688,641 and 5,677,125.

A genetic predisposition to disease in a human can also be detected by comparing expression levels of an mRNA or protein corresponding to a polynucleotide of the invention in a fetal tissue with levels associated in normal fetal tissue. Fetal tissues that are used for this purpose include, but are not limited to, amniotic fluid, chorionic villi, blood, and the blastomere of an *in vitro*-fertilized embryo. The comparable normal polynucleotide-related gene is obtained from any tissue. The mRNA or protein is obtained from a normal tissue of a human in which the polynucleotide-related gene is expressed. Differences such as alterations in the nucleotide sequence or size of the same product of the fetal polynucleotide-related gene or mRNA, or alterations in the molecular weight, amino acid sequence, or relative abundance of fetal protein, can indicate a germline mutation in the polynucleotide-related gene of the fetus, which indicates a genetic predisposition to disease. In general, diagnostic, prognostic, and other methods of the invention based on differential expression involve detection of a level or amount of a gene product, particularly a differentially expressed gene product, in a test sample obtained from a patient suspected of having or being susceptible to a disease (*e.g.*, breast cancer, lung cancer, colon cancer and/or metastatic forms thereof), and comparing the detected levels to those levels found in normal cells (*e.g.*, cells substantially unaffected by cancer) and/or other control cells (*e.g.*, to differentiate a cancerous cell from a cell affected by dysplasia). Furthermore, the severity of the disease can be assessed by comparing the detected levels of a differentially expressed gene product with those levels detected in samples representing the levels of differentially gene product associated with varying degrees of severity of disease. It should be noted that use of the term "diagnostic" herein is not necessarily meant to exclude "prognostic" or "prognosis," but rather is used as a matter of convenience.

The term "differentially expressed gene" is generally intended to encompass a polynucleotide that can, for example, include an open reading frame

encoding a gene product (*e.g.*, a polypeptide), and/or introns of such genes and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome. In general, a difference in expression level associated with a decrease in expression level of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% or more is indicative of a differentially expressed gene of interest, *i.e.*, a gene that is underexpressed or down-regulated in the test sample relative to a control sample. Furthermore, a difference in expression level associated with an increase in expression of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% and can be at least about 1 ½-fold, usually at least about 2-fold to about 10-fold, and can be about 100-fold to about 1,000-fold increase relative to a control sample is indicative of a differentially expressed gene of interest, *i.e.*, an overexpressed or up-regulated gene.

"Differentially expressed polynucleotide" as used herein means a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotides" is also meant to encompass fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

"Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, as well as to the prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy). The present invention particularly encompasses diagnosis of subjects in the context of breast cancer (*e.g.*, carcinoma *in situ* (*e.g.*, ductal carcinoma *in situ*), estrogen receptor (ER)-positive breast cancer, ER-negative breast cancer, or other forms and/or stages of breast cancer), lung cancer (*e.g.*, small cell carcinoma, non-small cell carcinoma, mesothelioma, and

other forms and/or stages of lung cancer), and colon cancer (*e.g.*, adenomatous polyp, colorectal carcinoma, and other forms and/or stages of colon cancer).

"Sample" or "biological sample" as used throughout here are generally meant to refer to samples of biological fluids or tissues, particularly samples obtained from tissues, especially from cells of the type associated with the disease for which the diagnostic application is designed (*e.g.*, ductal adenocarcinoma), and the like. "Samples" is also meant to encompass derivatives and fractions of such samples (*e.g.*, cell lysates). Where the sample is solid tissue, the cells of the tissue can be dissociated or tissue sections can be analyzed.

Methods of the subject invention useful in diagnosis or prognosis typically involve comparison of the abundance of a selected differentially expressed gene product in a sample of interest with that of a control to determine any relative differences in the expression of the gene product, where the difference can be measured qualitatively and/or quantitatively. Quantitation can be accomplished, for example, by comparing the level of expression product detected in the sample with the amounts of product present in a standard curve. A comparison can be made visually; by using a technique such as densitometry, with or without computerized assistance; by preparing a representative library of cDNA clones of mRNA isolated from a test sample, sequencing the clones in the library to determine that number of cDNA clones corresponding to the same gene product, and analyzing the number of clones corresponding to that same gene product relative to the number of clones of the same gene product in a control sample; or by using an array to detect relative levels of hybridization to a selected sequence or set of sequences, and comparing the hybridization pattern to that of a control. The differences in expression are then correlated with the presence or absence of an abnormal expression pattern. A variety of different methods for determining the nucleic acid abundance in a sample are known to those of skill in the art (see, *e.g.*, WO 97/27317).In general, diagnostic assays of the invention involve detection of a gene product of a the polynucleotide sequence (*e.g.*, mRNA or polypeptide) that corresponds to a sequence of SEQ ID NOs:1-3351. The patient from whom the sample is obtained can be apparently healthy, susceptible to disease (*e.g.*, as determined by family history or exposure to certain environmental factors), or can already be identified as having a condition in which altered expression of a gene product of the invention is implicated.

Diagnosis can be determined based on detected gene product expression levels of a gene product encoded by at least one, preferably at least two or more, at least

3 or more, or at least 4 or more of the polynucleotides having a sequence set forth in SEQ ID NOs:1-3351, and can involve detection of expression of genes corresponding to all of SEQ ID NOs:1-3351 and/or additional sequences that can serve as additional diagnostic markers and/or reference sequences. Where the diagnostic method is designed to detect the presence or susceptibility of a patient to cancer, the assay preferably involves detection of a gene product encoded by a gene corresponding to a polynucleotide that is differentially expressed in cancer. Examples of such differentially expressed polynucleotides are described in the Examples below. Given the provided polynucleotides and information regarding their relative expression levels provided herein, assays using such polynucleotides and detection of their expression levels in diagnosis and prognosis will be readily apparent to the ordinarily skilled artisan.

Any of a variety of detectable labels can be used in connection with the various embodiments of the diagnostic methods of the invention. Suitable detectable labels include fluorochromes, (*e.g.*, fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (*e.g.*, ^{32}P, ^{35}S, ^{3}H, *etc.*), and the like. The detectable label can involve a two stage systems (*e.g.*, biotin-avidin, hapten-anti-hapten antibody, *etc.*)

Reagents specific for the polynucleotides and polypeptides of the invention, such as antibodies and nucleotide probes, can be supplied in a kit for detecting the presence of an expression product in a biological sample. The kit can also contain buffers or labeling components, as well as instructions for using the reagents to detect and quantify expression products in the biological sample. Exemplary embodiments of the diagnostic methods of the invention are described below in more detail.

Polypeptide detection in diagnosis. In one embodiment, the test sample is assayed for the level of a differentially expressed polypeptide. Diagnosis can be accomplished using any of a number of methods to determine the absence or presence or altered amounts of the differentially expressed polypeptide in the test sample. For example, detection can utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells can be permeabilized to stain cytoplasmic molecules. In general, antibodies that specifically bind a differentially expressed polypeptide of the invention are added to a sample, and

incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody can be detectably labeled for direct detection (*e.g.*, using radioisotopes, enzymes, fluorescers, chemiluminescers, and the like), or can be used in conjunction with a second stage antibody or reagent to detect binding (*e.g.*, biotin with horseradish peroxidase-conjugated avidin, a secondary antibody conjugated to a fluorescent compound, *e.g.*, fluorescein, rhodamine, Texas red, *etc.*). The absence or presence of antibody binding can be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, *etc.* Any suitable alternative methods can of qualitative or quantitative detection of levels or amounts of differentially expressed polypeptide can be used, for example ELISA, western blot, immunoprecipitation, radioimmunoassay, *etc.*

mRNA detection. The diagnostic methods of the invention can also or alternatively involve detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotides of the invention. Any suitable qualitative or quantitative methods known in the art for detecting specific mRNAs can be used. mRNA can be detected by, for example, *in situ* hybridization in tissue sections, by reverse transcriptase-PCR, or in Northern blots containing poly A+ mRNA. One of skill in the art can readily use these methods to determine differences in the size or amount of mRNA transcripts between two samples. mRNA expression levels in a sample can also be determined by generation of a library of expressed sequence tags (ESTs) from the sample, where the EST library is representative of sequences present in the sample (Adams, et al., (1991) Science 252:1651). Enumeration of the relative representation of ESTs within the library can be used to approximate the relative representation of the gene transcript within the starting sample. The results of EST analysis of a test sample can then be compared to EST analysis of a reference sample to determine the relative expression levels of a selected polynucleotide, particularly a polynucleotide corresponding to one or more of the differentially expressed genes described herein. Alternatively, gene expression in a test sample can be performed using serial analysis of gene expression (SAGE) methodology (*e.g.*, Velculescu et al., *Science* (1995) *270*:484) or differential display (DD) methodology (see, *e.g.*, U.S. Patent NOs. 5,776,683 and 5,807,680).

Alternatively, gene expression can be analyzed using hybridization analysis. Oligonucleotides or cDNA can be used to selectively identify or capture DNA or RNA of specific sequence composition, and the amount of RNA or cDNA hybridized to a known capture sequence determined qualitatively or quantitatively, to provide

information about the relative representation of a particular message within the pool of cellular messages in a sample. Hybridization analysis can be designed to allow for concurrent screening of the relative expression of hundreds to thousands of genes by using, for example, array-based technologies having high density formats, including filters, microscope slides, or microchips, or solution-based technologies that use spectroscopic analysis (*e.g.*, mass spectrometry). One exemplary use of arrays in the diagnostic methods of the invention is described below in more detail.

Use of a single gene in diagnostic applications. The diagnostic methods of the invention can focus on the expression of a single differentially expressed gene. For example, the diagnostic method can involve detecting a differentially expressed gene, or a polymorphism of such a gene (*e.g.*, a polymorphism in an coding region or control region), that is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level and/or affect activity of the encoded protein, *etc.*

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, *e.g.*, a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express a differentially expressed gene can be used as a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis, and a detectable label can be included in the amplification reaction (*e.g.*, using a detectably labeled primer or detectably labeled oligonucleotides) to facilitate detection. Alternatively, various methods are also known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, see *e.g.*, Riley et al., *Nucl. Acids Res.* (1990) *18*:2887; and Delahunty et al., *Am. J. Hum. Genet.* (1996) *58*:1239.

The amplified or cloned sample nucleic acid can be analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy or other methods, and the sequence of bases compared to a selected sequence, *e.g.*, to a wild-type sequence. Hybridization with the polymorphic or variant sequence can also be used to determine its presence in a sample (*e.g.*, by Southern blot, dot blot, *etc.*). The hybridization pattern of a polymorphic or variant sequence and a control sequence to an array of oligonucleotide probes immobilized on a solid support, as described in U.S. Patent No. 5,445,934, or in WO 95/35505, can also be used as a means of identifying

polymorphic or variant sequences associated with disease. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in a gene can be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that can affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in proteins can be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded protein can be determined by comparison with the wild-type protein.

Pattern matching in diagnosis using arrays. In another embodiment, the diagnostic and/or prognostic methods of the invention involve detection of expression of a selected set of genes in a test sample to produce a test expression pattern (TEP). The TEP is compared to a reference expression pattern (REP), which is generated by detection of expression of the selected set of genes in a reference sample (*e.g.*, a positive or negative control sample). The selected set of genes includes at least one of the genes of the invention, which genes correspond to the polynucleotide sequences of SEQ ID NOs:1-3351. Of particular interest is a selected set of genes that includes genes differentially expressed in the disease for which the test sample is to be screened.

"Reference sequences" or "reference polynucleotides" as used herein in the context of differential gene expression analysis and diagnosis/prognosis refers to a selected set of polynucleotides, which selected set includes at least one or more of the differentially expressed polynucleotides described herein. A plurality of reference sequences, preferably comprising positive and negative control sequences, can be included as reference sequences. Additional suitable reference sequences are found in Genbank, Unigene, and other nucleotide sequence databases (including, *e.g.*, expressed sequence tag (EST), partial, and full-length sequences).

"Reference array" means an array having reference sequences for use in hybridization with a sample, where the reference sequences include all, at least one of,

or any subset of the differentially expressed polynucleotides described herein. Usually such an array will include at least 3 different reference sequences, and can include any one or all of the provided differentially expressed sequences. Arrays of interest can further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for screening for a disease or disorder (*e.g.*, cancer, dysplasia, or other related or unrelated diseases, disorders, or conditions). The oligonucleotide sequence on the array will usually be at least about 12 nt in length, and can be of about the length of the provided sequences, or can extend into the flanking regions to generate fragments of 100 nt to 200 nt in length or more. Reference arrays can be produced according to any suitable methods known in the art. For example, methods of producing large arrays of oligonucleotides are described in U.S. Patent NOs. 5,134,854 and 5,445,934 using light-directed synthesis techniques. Using a computer controlled system, a heterogeneous array of monomers is converted, through simultaneous coupling at a number of reaction sites, into a heterogeneous array of polymers. Alternatively, microarrays are generated by deposition of pre-synthesized oligonucleotides onto a solid substrate, for example as described in PCT published application no. WO 95/35505.

A "reference expression pattern" or "REP" as used herein refers to the relative levels of expression of a selected set of genes, particularly of differentially expressed genes, that is associated with a selected cell type, *e.g.*, a normal cell, a cancerous cell, a cell exposed to an environmental stimulus, and the like. A "test expression pattern" or "TEP" refers to relative levels of expression of a selected set of genes, particularly of differentially expressed genes, in a test sample (*e.g.*, a cell of unknown or suspected disease state, from which mRNA is isolated).

REPs can be generated in a variety of ways according to methods well known in the art. For example, REPs can be generated by hybridizing a control sample to an array having a selected set of polynucleotides (particularly a selected set of differentially expressed polynucleotides), acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the REP with a TEP. Alternatively, all expressed sequences in a control sample can be isolated and sequenced, *e.g.*, by isolating mRNA from a control sample, converting the mRNA into cDNA, and sequencing the cDNA. The resulting sequence information roughly or precisely reflects the identity and relative number of expressed sequences in the sample. The sequence information can then be stored in a format (*e.g.*, a computer-readable format) that allows for ready comparison of the REP with a TEP. The REP can be

normalized prior to or after data storage, and/or can be processed to selectively remove sequences of expressed genes that are of less interest or that might complicate analysis (*e.g.*, some or all of the sequences associated with housekeeping genes can be eliminated from REP data).

TEPs can be generated in a manner similar to REPs, *e.g.*, by hybridizing a test sample to an array having a selected set of polynucleotides, particularly a selected set of differentially expressed polynucleotides, acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the TEP with a REP. The REP and TEP to be used in a comparison can be generated simultaneously, or the TEP can be compared to previously generated and stored REPs.

In one embodiment of the invention, comparison of a TEP with a REP involves hybridizing a test sample with a reference array, where the reference array has one or more reference sequences for use in hybridization with a sample. The reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Hybridization data for the test sample is acquired, the data normalized, and the produced TEP compared with a REP generated using an array having the same or similar selected set of differentially expressed polynucleotides. Probes that correspond to sequences differentially expressed between the two samples will show decreased or increased hybridization efficiency for one of the samples relative to the other.

Methods for collection of data from hybridization of samples with a reference arrays are well known in the art. For example, the polynucleotides of the reference and test samples can be generated using a detectable fluorescent label, and hybridization of the polynucleotides in the samples detected by scanning the microarrays for the presence of the detectable label using, for example, a microscope and light source for directing light at a substrate. A photon counter detects fluorescence from the substrate, while an x-y translation stage varies the location of the substrate. A confocal detection device that can be used in the subject methods is described in U.S. Patent No. 5,631,734. A scanning laser microscope is described in Shalon et al., *Genome Res.* (1996) *6*:639. A scan, using the appropriate excitation line, is performed for each fluorophore used. The digital images generated from the scan are then combined for subsequent analysis. For any particular array element, the ratio of the fluorescent signal from one sample (*e.g.*, a test sample) is compared to the fluorescent signal from another sample (*e.g.*, a reference sample), and the relative signal intensity determined.

Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, *i.e.*, data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the targets from the remaining data. The resulting data can be displayed as an image with the intensity in each region varying according to the binding affinity between targets and probes.

In general, the test sample is classified as having a gene expression profile corresponding to that associated with a disease or non-disease state by comparing the TEP generated from the test sample to one or more REPs generated from reference samples (*e.g.*, from samples associated with cancer or specific stages of cancer, dysplasia, samples affected by a disease other than cancer, normal samples, *etc.*). The criteria for a match or a substantial match between a TEP and a REP include expression of the same or substantially the same set of reference genes, as well as expression of these reference genes at substantially the same levels (*e.g.*, no significant difference between the samples for a signal associated with a selected reference sequence after normalization of the samples, or at least no greater than about 25% to about 40% difference in signal strength for a given reference sequence. In general, a pattern match between a TEP and a REP includes a match in expression, preferably a match in qualitative or quantitative expression level, of at least one of, all or any subset of the differentially expressed genes of the invention.

Pattern matching can be performed manually, or can be performed using a computer program. Methods for preparation of substrate matrices (*e.g.*, arrays), design of oligonucleotides for use with such matrices, labeling of probes, hybridization conditions, scanning of hybridized matrices, and analysis of patterns generated, including comparison analysis, are described in, for example, U.S. Patent No. 5,800,992.

Diagnosis, Prognosis and Management of Cancer

The polynucleotides of the invention and their gene products are of particular interest as genetic or biochemical markers (*e.g.*, in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions. For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and

therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting and gene therapy. Determining expression of certain polynucleotides and comparison of a patients profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the polynucleotides of the invention are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

The polynucleotides of the invention can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. Furthermore, a polynucleotide of the invention identified as important for one type of cancer can also have implications for development or risk of development of other types of cancer, *e.g.*, where a polynucleotide is differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide that has clinical implications for metastatic colon cancer can also have clinical implications for stomach cancer or endometrial cancer.

Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes it is called Stage I. If it has spread only to the closest lymph nodes, it is called Stage II. In Stage III, the cancer has generally spread to the lymph nodes in near proximity to the site of the primary lesion. Cancers that have spread to a distant part of the body, such as the liver, bone, brain or other site, are Stage IV, the most advanced stage.

The polynucleotides of the invention can facilitate fine-tuning of the staging process by identifying markers for the aggresivity of a cancer, *e.g.*, the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive

therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

Grading of cancers. Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors being more aggressive than well differentiated or low-grade tumors. The following guidelines are generally used for grading tumors: 1) GX Grade cannot be assessed; 2) G1 Well differentiated; G2 Moderately well differentiated; 3) G3 Poorly differentiated; 4) G4 Undifferentiated. The polynucleotides of the invention can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressivity of a tumor, such as metastatic potential.

Detection of lung cancer. The polynucleotides of the invention can be used to detect lung cancer in a subject. Although there are more than a dozen different kinds of lung cancer, the two main types of lung cancer are small cell and nonsmall cell, which encompass about 90% of all lung cancer cases. Small cell carcinoma (also called oat cell carcinoma) usually starts in one of the larger bronchial tubes, grows fairly rapidly, and is likely to be large by the time of diagnosis. Nonsmall cell lung cancer (NSCLC) is made up of three general subtypes of lung cancer. Epidermoid carcinoma (also called squamous cell carcinoma) usually starts in one of the larger bronchial tubes and grows relatively slowly. The size of these tumors can range from very small to quite large. Adenocarcinoma starts growing near the outside surface of the lung and can vary in both size and growth rate. Some slowly growing adenocarcinomas are described as alveolar cell cancer. Large cell carcinoma starts near the surface of the lung, grows rapidly, and the growth is usually fairly large when diagnosed. Other less common forms of lung cancer are carcinoid, cylindroma, mucoepidermoid, and malignant mesothelioma.

The polynucleotides of the invention, *e.g.*, polynucleotides differentially expressed in normal cells versus cancerous lung cells (*e.g.*, tumor cells of high or low metastatic potential) or between types of cancerous lung cells (*e.g.*, high metastatic versus low metastatic), can be used to distinguish types of lung cancer as well as identifying traits specific to a certain patient's cancer and selecting an appropriate

therapy. For example, if the patient's biopsy expresses a polynucleotide that is associated with a low metastatic potential, it may justify leaving a larger portion of the patient's lung in surgery to remove the lesion. Alternatively, a smaller lesion with expression of a polynucleotide that is associated with high metastatic potential may justify a more radical removal of lung tissue and/or the surrounding lymph nodes, even if no metastasis can be identified through pathological examination.

Detection of breast cancer. The majority of breast cancers are adenocarcinomas subtypes, which can be summarized as follows: 1) ductal carcinoma *in situ* (DCIS), including comedocarcinoma; 2) infiltrating (or invasive) ductal carcinoma (IDC); 3) lobular carcinoma *in situ* (LCIS); 4) infiltrating (or invasive) lobular carcinoma (ILC); 5) inflammatory breast cancer; 6) medullary carcinoma; 7) mucinous carcinoma; 8) Paget's disease of the nipple; 9) Phyllodes tumor; and 10) tubular carcinoma.

The expression of polynucleotides of the invention can be used in the diagnosis and management of breast cancer, as well as to distinguish between types of breast cancer. Detection of breast cancer can be determined using expression levels of any of the appropriate polynucleotides of the invention, either alone or in combination. Determination of the aggressive nature and/or the metastatic potential of a breast cancer can also be determined by comparing levels of one or more polynucleotides of the invention and comparing levels of another sequence known to vary in cancerous tissue, *e.g.*, ER expression. In addition, development of breast cancer can be detected by examining the ratio of expression of a differentially expressed polynucleotide to the levels of steroid hormones (*e.g.*, testosterone or estrogen) or to other hormones (*e.g.*, growth hormone, insulin). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous breast tissue, to discriminate between breast cancers with different cells of origin, to discriminate between breast cancers with different potential metastatic rates, *etc.*

Detection of colon cancer. The polynucleotides of the invention exhibiting the appropriate expression pattern can be used to detect colon cancer in a subject. Colorectal cancer is one of the most common neoplasms in humans and perhaps the most frequent form of hereditary neoplasia. Prevention and early detection are key factors in controlling and curing colorectal cancer. Colorectal cancer begins as polyps, which are small, benign growths of cells that form on the inner lining of the colon. Over a period of several years, some of these polyps accumulate additional mutations and become cancerous. Multiple familial colorectal cancer disorders have

been identified, which are summarized as follows: 1) Familial adenomatous polyposis (FAP); 2) Gardner's syndrome; 3) Hereditary nonpolyposis colon cancer (HNPCC); and 4) Familial colorectal cancer in Ashkenazi Jews. The expression of appropriate polynucleotides of the invention can be used in the diagnosis, prognosis and management of colorectal cancer. Detection of colon cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression. Determination of the aggressive nature and/or the metastatic potential of a colon cancer can be determined by comparing levels of one or more polynucleotides of the invention and comparing total levels of another sequence known to vary in cancerous tissue, *e.g.*, expression of p53, DCC ras, lor FAP (see, *e.g.*, Fearon ER, et al., *Cell* (1990) *61*(*5*):759; Hamilton SR et al., *Cancer* (1993) *72*:957; Bodmer W, et al., *Nat Genet.* (1994) *4*(*3*):217; Fearon ER, *Ann N Y Acad Sci.* (1995) *768*:101). For example, development of colon cancer can be detected by examining the ratio of any of the polynucleotides of the invention to the levels of oncogenes (*e.g.*, ras) or tumor suppressor genes (*e.g.*, FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous colon tissue, to discriminate between colon cancers with different cells of origin, to discriminate between colon cancers with different potential metastatic rates, *etc.*

Use of Polynucleotides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full length genes can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (see, *e.g.*, U.S. Patent No. 5,010,175, and WO 91/17823). Agonists or antagonists of the polypeptides if the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, *etc.* The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo,* that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a novel polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide of the invention, and at least one peptide agonist or antagonist of the novel binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the novel receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

Pharmaceutical Compositions and Therapeutic Uses

Pharmaceutical compositions of the invention can comprise polypeptides, antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers,

and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, *e.g.*, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in *Remington's Pharmaceutical Sciences* (Mack Pub. Co., New Jersey, 1991).

Delivery Methods. Once formulated, the compositions of the invention can be (1) administered directly to the subject (*e.g.*, as polynucleotide or polypeptides); or (2) delivered *ex vivo*, to cells derived from the subject (*e.g.*, as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, *e.g.*, subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.*, International Publication No. WO 93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Once a gene corresponding to a polynucleotide of the invention has been found to correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the disorder can be amenable to treatment by administration of a

therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (*e.g.*, antisense, ribozyme, *etc.*).

The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic compositions agents of the invention includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. Preferably, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide disclosed herein. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

Receptor-mediated targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis et al., *Trends Biotechnol.* (1993) *11*:202; Chiou et al., *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer* (J.A. Wolff, ed.) (1994); Wu et al., *J. Biol. Chem.* (1988) *263*:621; Wu et al., *J. Biol. Chem.* (1994) *269*:542; Zenke et al., *Proc. Natl. Acad. Sci.* (USA) (1990) *87*:3655; Wu et al., *J. Biol. Chem.* (1991) *266*:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 mg to about 2 mg, about 5 mg to about 500 mg, and about 20 mg to about 100 mg of DNA can also be used during a gene therapy protocol. Factors such as method of action (*e.g.*, for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will

affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect. For polynucleotide-related genes encoding polypeptides or proteins with anti-inflammatory activity, suitable use, doses, and administration are described in U.S. Patent No. 5,654,173.

The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, *Cancer Gene Therapy* (1994) *1*:51; Kimura, *Human Gene Therapy* (1994) *5*:845; Connelly, *Human Gene Therapy* (1995) *1*:185; and Kaplitt, *Nature Genetics* (1994) 6:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, *e.g.*, WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; U.S. Patent No. 5, 219,740; WO 93/11230; WO 93/10218; U.S. Patent No. 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (*e.g.*, Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, *e.g.*, WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus as described in Curiel, *Hum. Gene Ther.* (1992) *3*:147 can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, *e.g.*, Curiel, *Hum. Gene Ther.* (1992) *3*:147); ligand-linked DNA(see, *e.g.*, Wu, *J. Biol. Chem. 264*:16985 (1989)); eukaryotic cell delivery vehicles cells (see, *e.g.*, U.S. Patent No. 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction

methods are described in WO 90/11092 and U.S. Patent No. 5,580,859. Liposomes that can act as gene delivery vehicles are described in U.S. Patent No. 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, *Mol. Cell Biol. 14*:2411 (1994), and in Woffendin, *Proc. Natl. Acad. Sci.* (1994) *91*:1581.

Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin et al., *Proc. Natl. Acad. Sci. USA 91*(24):11581 (1994). Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation (see, *e.g.*, U.S. Patent No. 5,206,152 and WO 92/11033). Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun (see, *e.g.*, U.S. Patent No. 5,149,655); use of ionizing radiation for activating transferred gene (see, *e.g.*, U.S. Patent No. 5,206,152 and WO 92/11033).

The present invention will now be illustrated by reference to the following examples which set forth particularly advantageous embodiments. However, it should be noted that these embodiments are illustrative and are not to be construed as restricting the invention in any way.

EXAMPLES

EXAMPLE 1

SOURCE OF BIOLOGICAL MATERIALS AND OVERVIEW OF NOVEL POLYNUCLEOTIDES EXPRESSED BY THE BIOLOGICAL MATERIALS

Cell lines and human normal and tumor tissue were used to construct cDNA libraries from mRNA isolated from the cells and tissues. Most sequences were about 275-300 nucleotides in length. The cells lines include Km12L4-A cell line, a high metastatic colon cancer cell line (Morika, W. A. K. et al., *Cancer Research* (1988) *48*:6863). The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B2 surgical specimen (Morikawa et al. *Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman et al. *Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling et al. *Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as model cell lines for the study of colon cancer (see, *e.g.*, Moriakawa et al., *supra*; Radinsky et al. *Clin. Cancer Res.* (1995) 1:19; Yeatman et al., (1995) *supra*; Yeatman et al., *Clin. Exp. Metastasis* (1996) 14:246). These and other cell lines and tissue are described in Table 6.

The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol. 266*:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams et al., eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie et al. *Comput. Chem.* (1993) *17*:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, *e.g.*, Alu repeats. The sequences remaining after masking were then used in a BLASTN vs. Genbank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1 x 10^{-40} were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the Genbank search), (2) weak similarity (greater than 45% identity and p value of less than 1×10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1×10^{-5}). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1×10^{-40} were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1×10^{-40} were discarded. Sequences with a p value of less than 1×10^{-65} when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1×10^{-40}, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1×10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 3351 sequences listed in the accompanying Sequence Listing. Each identified polynucleotide represents sequence from at least a partial mRNA transcript. Polynucleotides that were determined to be novel were assigned a sequence identification number.

The novel polynucleotides were assigned sequence identification numbers SEQ ID NOs:1-3351. The first 1847 DNA sequences corresponding to the novel polynucleotides are provided in the Sequence Listing in Table 1. DNA sequences corresponding to the novel polynucleotides of SEQ ID NOs:1848-3351 are provided in the Sequence Listing in Table 2. The DNA sequences of Table 2, while numbered SEQ ID 1-1504, correspond to SEQ ID NOs:1848-3351 in the Sequence Listing, *e.g.*, Table 2 SEQ ID 1 is SEQ ID NO:1848, Table 2 SEQ ID 2 is SEQ ID NO:1849, *etc.* Each DNA sequence in Table 4 is uniquely identified by a number that is 1847 less than its SEQ ID NO in the Sequence Listing. Tables 1 and 2 provide: 1) the SEQ ID NO assigned to each sequence for use in the present specification or a corresponding number; 2) the sequence name used as an internal identifier of the sequence; 3) the name assigned to the clone from which the

sequence was isolated; and 4) the number of the cluster to which the sequence is assigned (Cluster ID; where the cluster ID is 0, the sequence was not assigned to any cluster).

Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOs: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

EXAMPLE 2

RESULTS OF PUBLIC DATABASE SEARCH TO IDENTIFY FUNCTION OF GENE PRODUCTS

SEQ ID NOs:1-3351 were translated in all three reading frames to determine the best alignment with the individual sequences. These amino acid sequences and nucleotide sequences are referred to, generally, as query sequences, which are aligned with the individual sequences. Query and individual sequences were aligned using the BLAST programs, available over the world wide web at http://www.ncbi.nlm.nih.gov/BLAST/. Again the sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above in Example 1.

Tables 3 and 4 (inserted before the claims) show the results of the alignments. Table 3 contains alignment information for SEQ ID NOs:1-1847 and Table 4 contains alignment information for SEQ ID NOs:1848-3351. The DNA sequences of Table 4, while numbered SEQ ID 1-1504, correspond to SEQ ID NOs:1848-3351. Each DNA sequence in Table 4 is uniquely identified by a number that is 1847 less than its SEQ ID NO. Tables 3 and 4 refer to each sequence by its SEQ ID NO or a corresponding number, the accession numbers and descriptions of nearest neighbors from the Genbank and Non-Redundant Protein searches, and the p values of the search results.

For each of SEQ ID NOs:1-1847, the best alignment to a protein or DNA sequence is included in Table 3, and the best alignment for each of SEQ ID NOs:1848-3351 is included in Table 4. The activity of the polypeptide encoded by SEQ ID NOs:1-3351 is the same or similar to the nearest neighbor reported in Table 3 or 4. The accession number of the nearest neighbor is reported, providing a reference to the activities exhibited by the nearest neighbor. The search program and database used for the alignment also are indicated as well as a calculation of the p value.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of SEQ ID NOs:1-3351. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of SEQ ID NOs:1-3351.

EXAMPLE 3

MEMBERS OF PROTEIN FAMILIES

The sequences (SEQ ID NOs:1-3351) were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 5). "Start" and "stop" in Table 3 indicate the position within the individual sequences that align with the query sequence having the indicated SEQ ID NO. The direction indicates the orientation of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits because, for example, the particular sequence contains overlapping profile regions, and/or the sequence contains two different functional domains. These profile hits are described in more detail below.

Ank Repeats (ANK). SEQ ID NOs:187, 1268, 1804, 1819, 1830, 1839, 2652, 3015 and 3267 represent polynucleotides encoding an Ank repeat-containing protein. The ankyrin motif is a 33 amino acid sequence named for the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breeden et al., *Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno et al., *Development* (1997) *124(21)*:4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (*Biochem J. 290*:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-

protein interactions (Bork, *Proteins* (1993) *17(4)*:363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr et al., *Current Op. Cell Biol.* (1992) *4*:496; Bennet et al., *J. Biol. Chem.* (1980) *255*:6424).

ATPases Associated with Various Cellular Activities (ATPases). Sequences within SEQ ID NOs:431, 639, 2135, 2684, 2859, 3197 and 3266 correspond to a sequence that encodes a novel member of the "ATPases Associated with diverse cellular Activities" (AAA) protein family. The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids that contains an ATP-binding site (Froehlich et al., *J. Cell Biol.* (1991) *114*:443; Erdmann et al., *Cell* (1991) *64*:499; Peters et al., *EMBO J.* (1990) *9*:1757; Kunau et al., *Biochimie* (1993) *75*:209-224; Confalonieri et al., *BioEssays* (1995) *17*:639; http://yeamob.pci.chemie.uni-tuebingen.de/AAA/Description.html). The proteins that belong to this family either contain one or two AAA domains. In general, the AAA domains in these proteins act as ATP-dependent protein clamps (Confalonieri et al. (1995) *BioEssays 17*:639). In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, there is a highly conserved region located in the central part of the domain which was used in the development of the signature pattern. The consensus pattern is: [LIVMT]-x-[LIVMT]-[LIVMF]-x-[GATMC]-[ST]-[NS]-x(4)-[LIVM]- D-x-A-[LIFA]-x-R.

Bromodomain (bromodomain). SEQ ID NO:1814 represents a polynucleotide encoding a polypeptide having a bromodomain region (Haynes et al., 1992, Nucleic Acids Res. 20:2693-2603, Tamkun et al., 1992, Cell 68:561-572, and Tamkun, 1995, Curr. Opin. Genet. Dev. 5:473-477), which is a conserved region of about 70 amino acids. The bromodomain is thought to be involved in protein-protein interactions and may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation. The consensus pattern, which spans a major part of the bromodomain, is: [STANVF]-x(2)-F-x(4)-[DNS]-x(5,7)-[DENQTF]-Y-[HFY]-x(2)- [LIVMFY]-x(3)-[LIVM]-x(4)-[LIVM]-x(6,8)-Y-x(12,13)-[LIVM]-x(2)-N-[SACF]-x(2)-[FY].

Basic Region Plus Leucine Zipper Transcription Factors (BZIP). SEQ ID NOs:410, 552, 768, 822, 836, 1288, 1365, 1454, 1540, 1549, 1556, 1557, 1563, 1622, 1630, 1704, 1808, 2363, 2424, 3147, 3152, 3158 and 3208 represent polynucleotides encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding

transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. The consensus pattern for this protein family is: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

EF Hand (EFhand). SEQ ID NOs:820, 1755 and 3285 correspond to polynucleotides encoding a novel protein in the family of EF-hand proteins. Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand (Kawasaki et al., *Protein. Prof.* (1995) *2*:305-490). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand). The consensus pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW].

Ets Domain (Ets_Nterm). SEQ ID NO:1811 represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., *Eur. J. Biochem.* (1993) *211*:718). The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence and comprises an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

G-Protein Alpha Subunit (G-alpha). SEQ ID NO:1846 represents a polynucleotide encoding a novel polypeptide of the G-protein alpha subunit family. Guanine nucleotide binding proteins (G-proteins) are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion channels and enzymes that vary the concentration of

second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit binds GTP and exhibits GTPase activity. G-protein alpha subunits are 350-400 amino acids in length and have molecular weights in the range 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals, and fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon et al., *Science* (1993) *252*:802). They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications can be important for membrane association and high- affinity interactions with other proteins.

Helicases conserved C-terminal domain (helicase_C). SEQ ID NOs:1496, 2826 and 2871 represent polynucleotides encoding novel members of the DEAD/H helicase family. A number of eukaryotic and prokaryotic proteins have been characterized (Schmid S.R., et al., *Mol. Microbiol.* (1992) *6*:283; Linder P., et al., *Nature* (1989) *337*:121; Wassarman D.A., et al., *Nature* (1991) *349*:463) on the basis of their structural similarity. All are involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman T.C., *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578 (Errata). One of these motifs, called the "D-E-A-D-box", represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be "D-E-A-H-box" proteins (Wassarman D.A., et al., *Nature* (1991) *349*:463; Harosh I., et al., *Nucleic Acids Res.* (1991) *19*:6331; Koonin E.V. et al., *J. Gen. Virol.* (1992) *73*:989. The following signature patterns are used to identify members of both subfamilies: 1) [LIVMF](2)-D-E-A-D-[RKEN]-x-[LIVMFYGSTN]; and 2) [GSAH]-x-[LIVMF](3)-D-E-[ALIV]-H-[NECR].

Homeobox domain (homeobox). SEQ ID NOs:1676, 1820 and 1821 represent polynucleotides encoding proteins having a homeobox domain. The homeobox is a protein domain of 60 amino acids (Gehring In: Guidebook to the Homeobox Genes, Duboule D., Ed., pp. 1-10, Oxford University Press, Oxford, (1994); Buerglin In: Guidebook to the Homeobox Genes, pp25-72, Oxford University Press, Oxford, (1994); Gehring, *Trends Biochem. Sci.* (1992) 17:277-280; Gehring et al., *Annu. Rev. Genet.* (1986) *20*:147-173; Schofield, *Trends Neurosci.* (1987) *10*:3-6) first

identified in a number of Drosophila homeotic and segmentation proteins. It is extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Several proteins that contain a homeobox domain play an important role in development. Most of these proteins are sequence-specific DNA-binding transcription factors. The homeobox domain is also very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion.

A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHttHHHHHHHHHxxxxxxxxxx
1                                                          60
```

The pattern detects homeobox sequences 24 residues long and spans positions 34 to 57 of the homeobox domain. The consensus pattern is as follows: [LIVMFYG]-[ASLVR]-x(2)-[LIVMSTACN]-x-[LIVM]-x(4)-[LIV]-[RKNQESTAIY]-[LIVFSTNKH]-W-[FYVC]-x-[NDQTAH]-x(5)-[RKNAIMW].

MAP kinase kinase (mkk). SEQ ID NOs:29, 31, 196, 3175, 3190 and 3281 represent novel members of the MAP kinase kinase family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKKs are essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida et al., *Trends Biochem Sci* (1993) *18*:128-31; Ruderman, *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran et al., *Oncogene* (1998) *17*:1447-55; Kiefer et al., *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Protein Kinase (protkinase). SEQ ID NOs:1157, 1478, 1496, 2286, 2969 and 3190 represent polynucleotides encoding protein kinases. Protein kinases catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks S.K., et al., *FASEB J.* (1995) 9:576; Hunter T., *Meth. Enzymol.* (1991) *200*:3; Hanks S.K., et al., *Meth. Enzymol.* (1991) *200*:38; Hanks S.K.,

Curr. Opin. Struct. Biol. (1991) *1*:369; Hanks S.K. et al., *Science* (1988) *241*:42) are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme (Knighton D.R. et al., *Science* (1991) *253*:407). The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks S.K. et al., *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

The consensus patterns are as follows: 1) [LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K, where K binds ATP; 2) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3), where D is an active site residue; and 3) [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC], where D is an active site residue.

If a protein analyzed includes two of the above protein kinase signatures, the probability of it being a protein kinase is close to 100%.

Ras family proteins (ras). SEQ ID NOs:1688 and 3258 represent polynucleotides encoding novel members of the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, Biochemistry 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg2+ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, Biochemistry 30:4637-4648. A major consensus pattern of ras proteins is: D-T-A-G-Q-E-K-[LF]-G-G-L-R-[DE]-G-Y-Y.

Thioredoxin family active site (Thioredox). SEQ ID NO:1677 represents a polynucleotide encoding a protein having a thioredoxin family active site. Thioredoxins (Holmgren A., *Annu. Rev. Biochem.* (1985) *54*:237; Gleason F.K. et al., *FEMS Microbiol. Rev.* (1988) *54*:271; Holmgren, *A. J. Biol. Chem.* (1989) *264*:13963; Eklund H. et al., *Proteins* (1991) *11*:13) are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved. All PDI contains two or three (ERp72) copies of the thioredoxin domain. The consensus pattern is: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT] (where the two C's form the redox-active bond).

Trypsin (trypsin). SEQ ID NO:1410 corresponds to a novel serine protease of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* (1988) *334*:528). The consensus patterns for this trypsin protein family are: 1) [LIVM]-[ST]-A-[STAG]-H-C, where H is the active site residue; and 2) [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH], where S is the active site residue. All sequences known to belong to this family are detected by the above consensus sequences, except for 18 different proteases which have lost the first conserved glycine. If a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100%.

WD Domain, G-Beta Repeats (WD_domain). SEQ ID NOs:1336, 1380, 1711, 1762, 1909, 2218, 3047, 3108 and 3292 represent novel members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane

anchoring and receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). The consensus pattern for the WD domain/G-Beta repeat family is: [LIVMSTAC]-[LIVMFYWSTAGC]-[LIMSTAG]-[LIVMSTAGC]-x(2)-[DN]-x(2)-[LIVMWSTAC]-x-[LIVMFSTAG]-W-[DEN]-[LIVMFSTAGCN].

wnt Family of Developmental Signaling Proteins (Wnt_dev_sign). SEQ ID NO:1538 corresponds to a novel member of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse R., *Trends Genet.* (1988) *4*:291) is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). All wnt family proteins share the following features characteristics of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters. The consensus pattern, which is based upon a highly conserved region including three cysteines, is as follows: C-K-C-H-G-[LIVMT]-S-G-x-C.

Protein Tyrosine Phosphatase (Y_phosphatase). SEQ ID NO:1417 represents a polynucleotide encoding a protein tyrosine kinase. Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) (Fischer et al., *Science* (1991) *253*:401; Charbonneau et al., *Annu. Rev. Cell Biol.* (1992) *8*:463; Trowbridge, *J. Biol. Chem.* (1991) *266*:23517; Tonks et al., *Trends Biochem. Sci.* (1989) *14*:497; and Hunter, *Cell* (1989) *58*:1013) catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). Structurally, all known receptor PTPases are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. PTPase domains consist of about 300 amino acids. The search of two conserved cysteines has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important. The consensus pattern for PTPases is: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY]; C is the active site residue.

Zinc Finger, C2H2 Type (Zincfing C2H2). SEQ ID NOs:308, 807, 1324, 1503, 1527, 3081, 3193 and 3306 correspond to polynucleotides encoding novel members of the of the C2H2 type zinc finger protein family. Zinc finger domains (Klug et al., *Trends Biochem. Sci.* (1987) *12*:464; Evans et al., *Cell* (1988) *52*:1; Payre et al., *FEBS Lett.* (1988) *234*:245; Miller et al., *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci.* USA (1988) *85*:99) are nucleic acid-binding protein structures. In addition to the conserved zinc ligand residues, it has been shown that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld et al., *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue. The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.

Src homology 2. SEQ ID NOs:186, 2591, 3307 and 3339 represent polynucleotides encoding novel members of the family of Src homology 2 (SH2) proteins. The Src homology 2 (SH2) domain is a protein domain of about 100 amino acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps (Sadowski I. et al., *Mol. Cell. Biol. 6*:4396-4408 (1986)). Similar sequences are found in many other intracellular signal-transducing proteins (Russel R.B. et al., *FEBS Lett. 304*:15-20 (1992)). SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (Marangere L.E.M., Pawson T., *J. Cell Sci. Suppl. 18*:97-104 (1994); Pawson T., Schlessinger J., *Curr. Biol. 3*:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell. Biol. 3*:8-13 (1993); Pawson T., *Nature 373*:573-580 (1995)).

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol. 3*:828-837(1993)). The profile to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

Src homology 3. SEQ ID NO:234, 1832, and 1835 represent polynucleotides encoding novel members of the family of Src homology 3 (SH3) proteins. The Src homology 3 (SH3) domain is a small protein domain of about 60 amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (*e.g.*, Src, Abl, Lck) (Mayer B.J. et al.,

Nature 332:272-275 (1988)). Since then, it has been found in a great variety of other intracellular or membrane-associated proteins (Musacchio A. et al., *FEBS Lett. 307*:55-61 (1992); Pawson T., Schlessinger J., Curr. Biol. 3:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell Biol. 3*:8-13 (1993); Pawson T., *Nature 373*:573-580 (1995)).

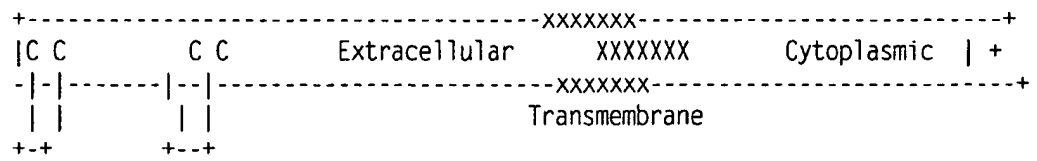
The SH3 domain has a characteristic fold which consists of five or six beta strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol. 3*:828-837 (1993)).

The function of the SH3 domain may be to mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton C.J., Campbell I.D., *Curr. Biol. 4*:615-617 (1994)).

In general SH3 domains are found as single copies in a given protein, but there are a significant number of proteins with two SH3 domains and a few with 3 or 4 copies.

Fibronectin type III. SEQ ID NOs:746 and 1192 represent polynucleotides encoding novel members of the family of fibronectin type III proteins. A number of receptors for lymphokines, hematopoeitic growth factors and growth hormone-related molecules have been found to share a common binding domain. (Bazan J.F., *Biochem. Biophys. Res. Commun. 164*:788-795 (1989); Bazan J.F., *Proc. Natl. Acad. Sci. U.S.A. 87*:6934-6938 (1990); Cosman D. et al., *Trends Biochem. Sci. 15*:265-270 (1990); d'Andrea A.D., Fasman G.D., Lodish H.F., *Cell 58*:1023-1024 (1989); d'Andrea A.D., Fasman G.D., Lodish H.F., *Curr. Opin. Cell Biol. 2*:648-651 (1990)).

The conserved region constitutes all or part of the extracellular ligand-binding region and is about 200 amino acid residues long. In the N-terminal of this domain there are two pairs of cysteines known, in the growth hormone receptor, to be involved in disulfide bonds.



Two patterns detect this family of receptors. The first one is derived from the first N-terminal disulfide loop, the second is a tryptophan-rich pattern located at the C-terminal extremity of the extracellular region.

A consensus for this protein family is: C-[LVFYR]-x(7,8)-[STIVDN]-C-x-W (The two C's are linked by a disulfide bond]. A second consensus for this protein family is: [STGL]-x-W-[SG]-x-W-S.

LIM domain containing proteins. SEQ ID NOs:1269, 1309, 1360, and 1386 represent polynucleotides encoding novel members of the family of LIM domain containing proteins. A number of proteins contain a conserved cysteine-rich domain of about 60 amino-acid residues. (Freyd G. et al., *Nature 344*:876-879 (1990); Baltz R. et al., *Plant Cell 4*:1465-1466 (1992); Sanchez-Garcia I., Rabbitts T.H., *Trends Genet. 10*:315-320 (1994)).

In the LIM domain, there are seven conserved cysteine residues and a histidine. The arrangement followed by these conserved residues is C-x(2)-C- x(16,23)-H-x(2)-[CH]-x(2)-C-x(2)-C-x(16,21)-C-x(2,3)-[CHD]. The LIM domain binds two zinc ions (Michelsen J.W. et al., *Proc. Natl. Acad. Sci. U.S.A. 90*:4404-4408 (1993)). LIM does not bind DNA, rather it seems to act as interface for protein-protein interaction. The consensus for this protein family is: C-x(2)-C-x(15,21)-[FYWH]-H-x(2)-[CH]-x(2)-C-x(2)-C-x(3)- [LIVMF]. The 5 C's and the H bind zinc.

C2 domain (protein kinase C like). SEQ ID NOs:1325 and 2282 represent polynucleotides encoding novel members of the family of C2 domain containing proteins. Some isozymes of protein kinase C (PKC) contain a domain, known as C2, of about 116 amino-acid residues, which is located between the two copies of the C1 domain (that bind phorbol esters and diacylglycerol) and the protein kinase catalytic domain. (Azzi A. et al., *Eur. J. Biochem. 208*:547-557 (1992); Stabel S., Semin. Cancer Biol. 5:277-284 (1994)).

The C2 domain is involved in calcium-dependent phospholipid binding (Davletov B.A., Suedhof T.C., *J. Biol. Chem. 268*:26386-26390 (1993)). Since domains related to the C2 domain are also found in proteins that do not bind calcium, other putative functions for the C2 domain include binding to inositol-1,3,5-tetraphosphate. (Fukuda M., et al., *J. Biol. Chem. 269*:29206-29211 (1994).)

The consensus pattern for the C2 domain is located in a conserved part of that domain, the connecting loop between beta strands 2 and 3. The profile for the C2 domain covers the total domain. The consensus for this protein family is:: [ACG]-x(2)-L-x(2,3)-D-x(1,2)-[NGSTLIF]-[GTMR]-x-[STAP]-D- [PA]-[FY]

Serine proteases, trypsin family, active sites. SEQ ID NO:1410 represents a polynucleotide encoding a novel member of the family of serine protease, trypsin proteins. The catalytic activity of the serine proteases from the trypsin family is

provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature 334*:528-530 (1988)).

A consensus for this protein family is: [LIVM]-[ST]-A-[STAG]-H-C [H is the active site residue]. A second consensus for this protein family is: [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]- [LIVMFYWH]-[LIVMFYSTANQH] [S is the active site residue].

RNA Recognition Motif Domain (RRM, RBD, or RNP). SEQ ID NOs: 1464 and 1514 represent polynucleotides encoding novel members of the family of RNA recognition motif domain proteins (Bandziulis R.J. et al., *Genes Dev. 3*:431-437 (1989); Dreyfuss G. et al., *Trends Biochem. Sci. 13*:86-91 (1988)).

Inside the putative RNA-binding domain there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif); the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation:

```
xxxxxxx######xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx#######xxxxxxxxxxxxxxxxxxxxxxxx
      RNP-2                                  RNP-1
```

As a consensus pattern for this type of domain the RNP-1 motif was used. The consensus for this protein family is: [RK]-G-{EDRKHPCG}-[AGSCI]-[FY]-[LIVA]-x-[FYLM]

Phosphatidylinositol-specific phospholipase C, Y Domain. SEQ ID NO: 1707 represents a polynucleotide encoding a novel member of the phosphatidylinositol-specific phospholipase C, Y domain family of proteins. Phosphatidylinositol-specific phospholipase C (EC3.1.4.11), a eukaryotic intracellular enzyme, plays an important role in signal transduction processes (Meldrum E. et al., *Biochim. Biophys. Acta 1092*:49-71 (1991)). It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5- triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins (Rhee S.G., Choi K.D., *Adv. Second Messenger Phosphoprotein Res. 26*:35-61 (1992); Rhee S.G., Choi K.D., *J. Biol. Chem. 267*:12393-12396 (1992); Sternweis P.C., Smrcka A.V., Trends Biochem. *Sci. 17*:502-506 (1992)).

All eukaryotic PI-PLCs contain two regions of homology, referred to as "X-box" and "Y-box". The order of these two regions is the same (NH2-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain possibly involved in Ca-dependent membrane attachment.

Serine Carboxypeptidases. SEQ ID NO:1744 represents a polynucleotide encoding a novel member of the serine carboxypeptidases family of proteins. Carboxypeptidases may be either metallo carboxypeptidases or serine carboxypeptidases (EC 3.4.16.5 and EC 3.4.16.6). The catalytic activity of the serine carboxypeptidases, like that of the trypsin family serine proteases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine (Liao D.I., Remington S.J., *J. Biol. Chem. 265*:6528-6531 (1990)).

The sequences surrounding the active site serine and histidine residues are highly conserved in all these serine carboxypeptidases. A consensus for this protein family is: [LIVM]-x-[GTA]-E-S-Y-[AG]-[GS] [S is the active site residue]. A second consensus for this protein family is: [LIVF]-x(2)-[LIVSTA]-x-[IVPST]-x-[GSDNQL]-[SAGV]-[SG]-H-x- [IVAQ]-P-x(3)-[PSA] [H is the active site residue].

dsrm Double-Stranded RNA Binding Motif. SEQ ID NO:1818 represents a polynucleotide encoding a novel member of the dsrm double-stranded RNA binding motif proteins. In eukaryotic cells, a multitude of RNA-binding proteins play key roles in the posttranscriptional regulation of gene expression. Characterization of these proteins has led to the identification of several RNA-binding motifs. Several human and other vertebrate genetic disorders are caused by aberrant expression of RNA-binding proteins. (C. G. Burd & G. Dreyfuss, *Science 265*: 615-621 (1994)).

Proteins containing double stranded RNA binding motifs bind to specific RNA targets. Double stranded RNA binding motifs are exemplified by interferon-induced protein kinase in humans, which is part of the cellular response to dsRNA.

SEQ ID NOs:2577, 3183 and 3195 encode members of the 4 trans-membrane integral membrane protein family. This family consists of type III proteins, which are integral membrane proteins that contain a N-terminal membrane-anchoring domain that is not cleaved during biosynthesis, and which functions as a translocation

signal and a membrane anchor. The proteins also have three additional transmembrane regions. The consensus pattern is: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF] (2)-G-C-x-[GA]-[STA]-x(20-[eG]-x(20-[CwN]-[LIVM](2).

SEQ ID NO:2944 encodes a polypeptide having a calpain large subunit, domain III. Calpains are a family of intracellular proteases that play a variety of biological roles. Calpain 3, also known as p94, is predominantly expressed in skeletal muscle and plays a role in limb-girdle muscular dystrophy type 2A. (Sorimachi, H. et al., Biochem. J. 328:721-732, 1997).

SEQ ID NOs:1911 and 1980 encode polypeptides having a C3HC4 type zinc finger domain (RING finger), which is a cysteine-rich domain of 40 to 60 residues that binds two atoms of zinc, and is believed to be involved in mediating protein-protein interactions. Mammalian proteins of this family include V(D)J recombination activating protein, which activates the rearrangement of immunoglobulin and T-cell receptor genes; breast cancer type 1 susceptibility protein (BRCA1); bmi-1 proto-oncogene; cbl proto-oncogene; and mel-18 protein, which is expressed in a variety of tumor cells and is a transcriptional repressor that recognizes and binds a specific DNA sequence. The consensus pattern is: C-x-H-x-[LIVMFY]-C-x(2)-C-[LIVMYA].

SEQ ID NO:3274 encodes a eukaryotic transcription factor with a fork head domain, of about 100 amino acid residues. Proteins of this group are transcription factors, including mammalian transcription factors HNF-3-alpha, -beta, and -gamma; interleukin-enhancer binding factor; and HTLF, which binds to a region of human T-cell leukemia virus long terminal repeat. The consensus pattern is [KR]-P-[PTQ]-[FYLVQH]-S-[FY]x(2)-[LIVM]-X(3,4)-[AC]-[LIM].

SEQ ID NO:3345 encodes a polypeptide having a PDZ domain. Several dozen signaling proteins belong to this group of proteins that have 80-100 residue repeats known as PDZ domains. Several of the proteins interact with the C-terminal tetrapeptide motifs X-Ser/Thr/X-Val-COO- of ion channels and/or receptors. (Ponting, C. P., Protein Sci. 6;464-468, 1997.)

SEQ ID NO:3351 encodes a polypeptide in the family of phorbol esters/glycerol binding proteins. Phorbol esters (PE) are analogues of diacylglycerol (DAG) and potent tumor promoters. DAG activates a family of serine-threonine protein kinases, known as protein kinase C. The N-terminal region of protein kinase C binds PE and DAG, and contains one or two copies of a cysteine-rich domain of about 50 amino acid residues. Other proteins having this domain include diacylglycerol kinase; the vav oncogene; and N-chimaerin, a brain-specific protein. The DAG/PE binding

domain binds two zinc ions through the six cysteines and two histidines that are conserved in the domain. The consensus pattern is: H-x-[LIVMFYW]-x(8, 11)-C-x(2)-C-x-(3)-[LIVMFC]-x(5, 10)-C-x(2)-C-x(4)-[HD]-x(2)-C-x(5, 9)-C.

SEQ ID NO:2216 encodes a polypeptide having a WW/rsp5/WWP domain. The protein is named for the presence of conserved aromatic positions, generally tryptophan, as well as a conserved proline. Proteins having the domain include dystrophin, vertebrate YAP protein, and IQGAP, a human GTPase activating protein which acts on ras. The consensus pattern is: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

SEQ ID NO:2428 encodes a member of the dual specificity phosphatase family, having a catalytic domain, and SEQ IDS NOs:2281 and 2310 encode members of the protein tyrosine phosphatase family. These families are related and classified as tyrosine specific protein phosphatases. The enzymes catalyze the removal of a phosphate group from a tyrosine residue, and are important in the control of cell growth, proliferation, differentiation, and transformation. The consensus pattern is [LIVMF]-H-C-x(2)-G-x-(3)-[STC]-[STAGP]-x-[LIVMFY].

Table 1

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
1	377044	RTA00002676F.p.11.2.P.Seq	F	M000393294.C01	CH09LNL
2	377708	RTA00002683F.m.01.2.P.Seq	F	M00040089A:G08	CH09LNL
3	427782	RTA00002666F.l.06.1.P.Seq	F	M00032638D:A06	CH08LNH
4	29372	RTA00002712F.a.06.1.P.Seq	F	M00023282A:C02	CH04MAL
5	455003	RTA00002694F.b.02.1.P.Seq	F	M00043419D:A10	CH20COHLV
6	380625	RTA00002684F.d.03.2.P.Seq	F	M00040118D:G10	CH09LNL
7	450959	RTA00002691F.b.05.3.P.Seq	F	M00043306D:B07	CH17COHLV
8	397851	RTA00002680F.b.04.1.P.Seq	F	M00039775A:A09	CH09LNL
9	20652	RTA00002710F.k.01.1.P.Seq	F	M00022440B:E01	CH03MAH
10	97830	RTA00002663F.k.18.1.P.Seq	F	M00022767B:G11	CH03MAH
11	373071	RTA00002670F.j.23.1.P.Seq	F	M00033442A:D06	CH09LNL
12	162369	RTA00002713F.e.01.1.P.Seq	F	M00027292D:F10	CH04MAL
13	401247	RTA00002685F.f.15.2.P.Seq	F	M00039508A:C12	CH12EDT
14	430738	RTA00002669F.i.15.3.P.Seq	F	M00033231D:B09	CH08LNH
15	46779	RTA00002711F.c.14.1.P.Seq	F	M00022860C:G04	CH03MAH
16	375772	RTA00002681F.p.01.2.P.Seq	F	M00039909C:G05	CH09LNL
17	430689	RTA00002669F.j.01.3.P.Seq	F	M00033243B:A05	CH08LNH
18	376546	RTA00002677F.d.07.2.P.Seq	F	M00039345C:C12	CH09LNL
19	430041	RTA00002667F.f.17.1.P.Seq	F	M00032790B:A07	CH08LNH
20	431643	RTA00002669F.l.16.1.P.Seq	F	M00033276D:H09	CH08LNH
21	19422	RTA00002709F.c.02.1.P.Seq	F	M00005449B:B10	CH02COH
22	376802	RTA00002677F.c.18.2.P.Seq	F	M00039344B:G07	CH09LNL
23	376814	RTA00002674F.h.02.1.P.Seq	F	M00039139C:G12	CH09LNL
24	375492	RTA00002677F.m.19.2.P.Seq	F	M00039418B:D08	CH09LNL
25	379114	RTA00002681F.n.24.2.P.Seq	F	M00039903C:F03	CH09LNL
26	380668	RTA00002670F.p.11.1.P.Seq	F	M00033581C:H10	CH09LNL
27	213817	RTA00002664F.i.19.2.P.Seq	F	M00027634A:D11	CH04MAL
28	375749	RTA00002680F.f.23.1.P.Seq	F	M00039795D:G06	CH09LNL
29	430896	RTA00002669F.b.20.4.P.Seq	F	M00033185C:D01	CH08LNH
30	380462	RTA00002670F.o.01.1.P.Seq	F	M00033570B:E06	CH09LNL
31	430896	RTA00002669F.b.20.3.P.Seq	F	M00033185C:D01	CH08LNH
32	376996	RTA00002676F.p.13.2.P.Seq	F	M00039329C:B10	CH09LNL
33	374846	RTA00002677F.k.19.2.P.Seq	F	M00039412D:G06	CH09LNL
34	379075	RTA00002672F.n.13.2.P.Seq	F	M00039039B:E03	CH09LNL
35	374172	RTA00002673F.k.16.2.P.Seq	F	M00039097D:D06	CH09LNL
36	373104	RTA00002683F.o.15.2.P.Seq	F	M00040098D:G12	CH09LNL
37	186302	RTA00002713F.m.21.1.P.Seq	F	M00027591B:C04	CH04MAL
38	427947	RTA00002665F.o.01.1.P.Seq	F	M00032495B:D02	CH08LNH
39	375180	RTA00002673F.d.17.1.P.Seq	F	M00039064D:H09	CH09LNL
40	377584	RTA00002683F.l.22.2.P.Seq	F	M00040088C:E10	CH09LNL
41	377364	RTA00002678F.a.15.2.P.Seq	F	M00039432C:A01	CH09LNL
42	376347	RTA00002675F.l.08.1.P.Seq	F	M00039249C:G11	CH09LNL
43	446747	RTA00002689F.d.16.2.P.Seq	F	M00042740A:E09	CH15CON
44	28092	RTA00002711F.g.12.1.P.Seq	F	M00023032A:B05	CH03MAH
45	378206	RTA00002671F.a.20.3.P.Seq	F	M00033588C:G04	CH09LNL
46	378206	RTA00002671F.a.20.2.P.Seq	F	M00033588C:G04	CH09LNL
47	14940	RTA00002709F.d.11.1.P.Seq	F	M00005623A:G02	CH02COH

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
142	24352	RTA00002709F.a.05.1.P.Seq	F	M00004839C:H02	CH02COH
143	24354	RTA00002709F.a.03.1.P.Seq	F	M00004832D:H02	CH02COH
144	379114	RTA00002681F.o.01.2.P.Seq	F	M00039903C:F03	CH09LNL
145	19609	RTA00002709F.c.05.1.P.Seq	F	M00005457C:A03	CH02COH
146	21685	RTA00002709F.e.23.1.P.Seq	F	M00006581D:F08	CH02COH
147	380085	RTA00002682F.i.10.1.P.Seq	F	M00039987A:F09	CH09LNL
148	20700	RTA00002710F.i.18.1.P.Seq	F	M00022373A:B05	CH03MAH
149	379981	RTA00002682F.i.18.1.P.Seq	F	M00039988A:E06	CH09LNL
150	376591	RTA00002675F.c.01.1.P.Seq	F	M00039213A:D01	CH09LNL
151	92058	RTA00002663F.m.04.1.P.Seq	F	M00022895A:H08	CH03MAH
152	196936	RTA00002663F.m.02.1.P.Seq	F	M00022885C:H05	CH03MAH
153	430702	RTA00002668F.h.04.1.P.Seq	F	M00032990B:A11	CH08LNH
154	378448	RTA00002680F.n.21.2.P.Seq	F	M00039832A:B12	CH09LNL
155	41606	RTA00002713F.e.10.1.P.Seq	F	M00027301A:G05	CH04MAL
156	213817	RTA00002664F.i.19.1.P.Seq	F	M00027634A:D11	CH04MAL
157	373464	RTA00002671F.l.13.1.P.Seq	F	M00038327A:C11	CH09LNL
158	379483	RTA00002679F.k.12.1.P.Seq	F	M00039700B:D02	CH09LNL
159	375796	RTA00002680F.f.17.1.P.Seq	F	M00039795B:H10	CH09LNL
160	375796	RTA00002680F.f.17.2.P.Seq	F	M00039795B:H10	CH09LNL
161	120485	RTA00002663F.b.12.1.P.Seq	F	M00021665B:F12	CH03MAH
162	374291	RTA00002673F.f.17.1.P.Seq	F	M00039072C:E02	CH09LNL
163	380513	RTA00002677F.p.15.2.P.Seq	F	M00039428C:E01	CH09LNL
164	379416	RTA00002683F.j.07.2.P.Seq	F	M00040077D:C11	CH09LNL
165	378178	RTA00002680F.l.13.1.P.Seq	F	M00039820A:F11	CH09LNL
166	427947	RTA00002665F.n.24.1.P.Seq	F	M00032495B:D02	CH08LNH
167	427269	RTA00002665F.d.03.3.P.Seq	F	M00028212C:B08	CH08LNH
168	20451	RTA00002710F.j.10.1.P.Seq	F	M00022391B:E01	CH03MAH
169	377003	RTA00002683F.g.09.2.P.Seq	F	M00040062B:B05	CH09LNL
170	427759	RTA00002665F.o.11.1.P.Seq	F	M00032499C:A01	CH08LNH
171	427549	RTA00002668F.k.13.1.P.Seq	F	M00033034C:A06	CH08LNH
172	373881	RTA00002672F.b.20.2.P.Seq	F	M00038638D:H03	CH09LNL
173	188215	RTA00002664F.f.13.2.P.Seq	F	M00027200A:F02	CH04MAL
174	379683	RTA00002681F.d.04.2.P.Seq	F	M00039857B:G10	CH09LNL
175	380652	RTA00002678F.d.12.2.P.Seq	F	M00039455D:H04	CH09LNL
176	378334	RTA00002679F.h.10.1.P.Seq	F	M00039682C:H11	CH09LNL
177	377930	RTA00002680F.g.14.1.P.Seq	F	M00039798B:B02	CH09LNL
178	378692	RTA00002680F.o.20.3.P.Seq	F	M00039835A:F07	CH09LNL
179	32279	RTA00002709F.d.23.1.P.Seq	F	M00005673B:B12	CH02COH
180	376379	RTA00002680F.c.15.1.P.Seq	F	M00039782A:H10	CH09LNL
181	375963	RTA00002675F.i.12.1.P.Seq	F	M00039238A:B12	CH09LNL
182	378683	RTA00002680F.a.14.2.P.Seq	F	M00039773D:A09	CH09LNL
183	374946	RTA00002673F.j.24.2.P.Seq	F	M00039096A:E07	CH09LNL
184	429583	RTA00002666F.g.10.1.P.Seq	F	M00032584A:H08	CH08LNH
185	28338	RTA00002711F.e.17.1.P.Seq	F	M00022930C:E02	CH03MAH
186	427970	RTA00002665F.j.13.1.P.Seq	F	M00031368A:E10	CH08LNH
187	379650	RTA00002683F.h.22.2.P.Seq	F	M00040072C:G09	CH09LNL
188	379661	RTA00002676F.c.05.2.P.Seq	F	M00039277D:G10	CH09LNL

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
753	455136	RTA00002694F.a.08.1.P.Seq	F	M00042595A:B01	CH20COHLV
754	379001	RTA00002683F.o.02.2.P.Seq	F	M00040097A:C12	CH09LNL
755	374763	RTA00002673F.p.21.1.P.Seq	F	M00039118B:C05	CH09LNL
756	402508	RTA00002686F.o.15.1.P.Seq	F	M00040281D:B01	CH13EDT
757	431370	RTA00002669F.m.04.3.P.Seq	F	M00033288B:D12	CH08LNH
758	380500	RTA00002670F.p.19.1.P.Seq	F	M00033583B:E06	CH09LNL
759	376743	RTA00002678F.e.22.2.P.Seq	F	M00039461A:F04	CH09LNL
760	191690	RTA00002673F.m.19.1.P.Seq	F	M00039107C:E04	CH09LNL
761	374264	RTA00002671F.p.21.2.P.Seq	F	M00038620B:E09	CH09LNL
762	373020	RTA00002671F.b.20.2.P.Seq	F	M00033595A:C11	CH09LNL
763	375231	RTA00002671F.m.20.2.P.Seq	F	M00038387B:A07	CH09LNL
764	16180	RTA00002709F.j.17.1.P.Seq	F	M00006977D:A03	CH02COH
765	379403	RTA00002683F.c.17.2.P.Seq	F	M00040041C:C09	CH09LNL
766	375382	RTA00002677F.d.24.2.P.Seq	F	M00039381D:C02	CH09LNL
767	379653	RTA00002683F.c.03.2.P.Seq	F	M00040038D:G04	CH09LNL
768	377858	RTA00002681F.e.14.2.P.Seq	F	M00039864A:A07	CH09LNL
769	430861	RTA00002668F.h.18.1.P.Seq	F	M00032995C:C05	CH08LNH
770	376128	RTA00002677F.a.11.2.P.Seq	F	M00039334B:E03	CH09LNL
771	375009	RTA00002676F.n.20.2.P.Seq	F	M00039322A:F04	CH09LNL
772	429816	RTA00002667F.n.22.1.P.Seq	F	M00032871D:E11	CH08LNH
773	375657	RTA00002681F.h.13.2.P.Seq	F	M00039877C:C03	CH09LNL
774	427889	RTA00002666F.b.14.1.P.Seq	F	M00032530D:C02	CH08LNH
775	376761	RTA00002677F.g.03.2.P.Seq	F	M00039391D:F08	CH09LNL
776	44025	RTA00002684F.b.24.2.P.Seq	F	M00040115B:A04	CH09LNL
777	44025	RTA00002684F.c.01.2.P.Seq	F	M00040115B:A04	CH09LNL
778	392524	RTA00002681F.p.04.2.P.Seq	F	M00039909D:C02	CH09LNL
779	427252	RTA00002665F.b.13.1.P.Seq	F	M00028185B:A06	CH08LNH
780	374927	RTA00002673F.e.12.1.P.Seq	F	M00039068C:E06	CH09LNL
781	378226	RTA00002680F.g.09.1.P.Seq	F	M00039797C:G05	CH09LNL
782	217964	RTA00002664F.g.08.2.P.Seq	F	M00027299B:B12	CH04MAL
783	376368	RTA00002677F.b.14.2.P.Seq	F	M00039339A:H07	CH09LNL
784	377719	RTA00002677F.j.11.2.P.Seq	F	M00039407B:G02	CH09LNL
785	378081	RTA00002677F.e.16.2.P.Seq	F	M00039384C:E02	CH09LNL
786	89267	RTA00002662F.b.01.2.P.Seq	F	M00005445D:B01	CH02COH
787	374927	RTA00002673F.e.12.2.P.Seq	F	M00039068C:E06	CH09LNL
788	279054	RTA00002667F.b.23.1.P.Seq	F	M00032731B:C10	CH08LNH
789	377283	RTA00002682F.m.19.1.P.Seq	F	M00040016C:H12	CH09LNL
790	45318	RTA00002710F.l.05.1.P.Seq	F	M00022533A:A08	CH03MAH
791	188292	RTA00002664F.e.23.2.P.Seq	F	M00027162B:F05	CH04MAL
792	378872	RTA00002683F.c.20.2.P.Seq	F	M00040042B:A10	CH09LNL
793	427252	RTA00002665F.b.13.3.P.Seq	F	M00028185B:A06	CH08LNH
794	380618	RTA00002673F.j.12.2.P.Seq	F	M00039084C:G07	CH09LNL
795	35646	RTA00002667F.g.16.1.P.Seq	F	M00032797B:G02	CH08LNH
796	46407	RTA00002665F.c.10.3.P.Seq	F	M00028196D:A03	CH08LNH
797	373720	RTA00002674F.c.04.1.P.Seq	F	M00039124C:F03	CH09LNL
798	429693	RTA00002668F.f.05.1.P.Seq	F	M00032944B:B02	CH08LNH
799	377108	RTA00002678F.p.04.2.P.Seq	F	M00039636C:D11	CH09LNL

1797	375714	RTA00002677F.m.13.2.P.Seq	F	M00039417C:G01	CH09LNL
1798	51564	RTA00002712F.d.23.1.P.Seq	F	M00023398B:D12	CH04MAL
1799	399551	RTA00002687F.f.13.2.P.Seq	F	M00040203D:H11	CH14EDT
1800	133512	RTA00002693F.e.24.2.P.Seq	F	M00043200A:H09	CH19COP
1801	375176	RTA00002675F.p.13.1.P.Seq	F	M00039266D:H04	CH09LNL
1802	375704	RTA00002676F.h.13.2.P.Seq	F	M00039300C:G04	CH09LNL
1803	399551	RTA00002687F.f.13.1.P.Seq	F	M00040203D:H11	CH14EDT
1804	403357	RTA00002687F.i.05.2.P.Seq	F	M00040298B:G02	CH14EDT
1805	34513	RTA00002709F.c.22.1.P.Seq	F	M00005555A:A10	CH02COH
1806	121871	RTA00002713F.a.09.1.P.Seq	F	M00027198B:B08	CH04MAL
1807	32095	RTA00002662F.d.15.2.P.Seq	F	M00007112C:B10	CH02COH
1808	403183	RTA00002687F.n.02.1.P.Seq	F	M00040332D:B05	CH14EDT
1809	168691	RTA00002663F.j.02.1.P.Seq	F	M00022615D:G05	CH03MAH
1810	430854	RTA00002668F.p.21.2.P.Seq	F	M00033173D:C01	CH08LNH
1811	377987	RTA00002679F.h.08.1.P.Seq	F	M00039682A:C08	CH09LNL
1812	428408	RTA00002665F.p.23.1.P.Seq	F	M00032513D:F01	CH08LNH
1813	375930	RTA00002677F.h.03.2.P.Seq	F	M00039396D:B04	CH09LNL
1814	28453	RTA00002711F.h.07.1.P.Seq	F	M00023094A:B11	CH03MAH
1815	119478	RTA00002686F.n.07.1.P.Seq	F	M00040271C:D08	CH13EDT
1816	403189	RTA00002687F.g.16.2.P.Seq	F	M00040217D:B07	CH14EDT
1817	129692	RTA00002679F.e.13.1.P.Seq	F	M00039673A:F09	CH09LNL
1818	86668	RTA00002664F.a.10.2.P.Seq	F	M00023352B:F03	CH04MAL
1819	403357	RTA00002687F.i.05.1.P.Seq	F	M00040298B:G02	CH14EDT
1820	373198	RTA00002670F.p.01.2.P.Seq	F	M00033578D:G02	CH09LNL
1821	373198	RTA00002670F.o.24.2.P.Seq	F	M00033578D:G02	CH09LNL
1822	25233	RTA00002711F.b.06.1.P.Seq	F	M00022823C:C01	CH03MAH
1823	403429	RTA00002687F.a.07.2.P.Seq	F	M00039746D:D11	CH14EDT
1824	417119	RTA00002686F.i.14.1.P.Seq	F	M00040222D:G02	CH13EDT
1825	376066	RTA00002680F.c.12.2.P.Seq	F	M00039781D:D10	CH09LNL
1826	403189	RTA00002687F.g.16.1.P.Seq	F	M00040217D:B07	CH14EDT
1827	403429	RTA00002687F.a.07.1.P.Seq	F	M00039746D:D11	CH14EDT
1828	430975	RTA00002669F.j.06.3.P.Seq	F	M00033246C:E08	CH08LNH
1829	427544	RTA00002665F.e.03.1.P.Seq	F	M00028354A:B12	CH08LNH
1830	401155	RTA00002685F.o.12.1.P.Seq	F	M00039630A:C08	CH12EDT
1831	377005	RTA00002682F.k.15.1.P.Seq	F	M00040005D:B07	CH09LNL
1832	379032	RTA00002683F.a.07.1.P.Seq	F	M00040032A:D09	CH09LNL
1833	400097	RTA00002685F.g.19.2.P.Seq	F	M00039521A:A02	CH12EDT
1834	383401	RTA00002670F.k.13.2.P.Seq	F	M00033450C:A02	CH09LNL
1835	379032	RTA00002683F.a.07.2.P.Seq	F	M00040032A:D09	CH09LNL
1836	429663	RTA00002667F.m.21.1.P.Seq	F	M00032864B:B09	CH08LNH
1837	374018	RTA00002672F.a.14.2.P.Seq	F	M00038632C:B09	CH09LNL
1838	375409	RTA00002678F.n.02.2.P.Seq	F	M00039616B:C01	CH09LNL
1839	401155	RTA00002685F.o.12.2.P.Seq	F	M00039630A:C08	CH12EDT
1840	13958	RTA00002711F.b.02.1.P.Seq	F	M00022817A:H02	CH03MAH
1841	38767	RTA00002687F.a.11.1.P.Seq	F	M00039748C:F11	CH14EDT
1842	29398	RTA00002663F.c.23.1.P.Seq	F	M00022015B:B07	CH03MAH
1843	12453	RTA00002709F.c.23.2.P.Seq	F	M00005556B:D02	CH02COH
1844	38767	RTA00002687F.a.11.2.P.Seq	F	M00039748C:F11	CH14EDT
1845	279885	RTA00002671F.f.05.2.P.Seq	F	M00038279C:A11	CH09LNL

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
1	10600	RTA00002891F.j.07.1.P.Seq	F	M000037S8B:D07	CH01COH
2	18827	RTA00002900F.p.12.1.P.Seq	F	M00005413D:A05	CH02COH
3	1759	RTA00002923F.f.23.1.P.Seq	F	M00039248C:A08	CH09LNL
4	10924	RTA00002907F.k.12.1.P.Seq	F	M00022224A:C07	CH03MAH
5	45331	RTA00002903F.l.10.1.P.Seq	F	M00007037D:D10	CH02COH
6	42233	RTA00002912F.g.24.1.P.Seq	F	M00027359B:A06	CH04MAL
7	7211	RTA00002909F.h.06.1.P.Seq	F	M00022634A:C07	CH03MAH
8	21395	RTA00002890F.k.16.1.P.Seq	F	M00001637D:C12	CH01COH
9	3093	RTA00002923F.e.03.1.P.Seq	F	M00039225A:D11	CH09LNL
10	15806	RTA00002894F.f.07.1.P.Seq	F	M00003991A:C11	CH01COH
11	19739	RTA00002896F.d.12.1.P.Seq	F	M00004147C:E01	CH01COH
12	140879	RTA00002905F.c.17.1.P.Seq	F	M00007985C:D08	CH03MAH
13	29706	RTA00002908F.l.22.1.P.Seq	F	M00022487B:A08	CH03MAH
14	109581	RTA00002918F.i.08.1.P.Seq	F	M00032908A:D08	CH08LNH
15	25009	RTA00002906F.k.11.1.P.Seq	F	M00022016B:F01	CH03MAH
16	8328	RTA00002888F.e.07.1.P.Seq	F	M00001451C:E10	CH01COH
17	15045	RTA00002887F.e.06.1.P.Seq	F	M00001393C:E08	CH01COH
18	21216	RTA00002898F.p.22.1.P.Seq	F	M00004416B:G10	CH01COH
19	185754	RTA00002912F.l.09.1.P.Seq	F	M00027506B:G01	CH04MAL
20	11881	RTA00002909F.h.10.1.P.Seq	F	M00022638A:D03	CH03MAH
21	185989	RTA00002910F.h.12.1.P.Seq	F	M00022924C:F04	CH03MAH
22	9667	RTA00002923F.a.03.1.P.Seq	F	M00039162D:C04	CH09LNL
23	15817	RTA00002903F.o.03.1.P.Seq	F	M00007103D:C02	CH02COH
24	10198	RTA00002923F.j.09.1.P.Seq	F	M00039294C:B09	CH09LNL
25	6355	RTA00002894F.p.12.1.P.Seq	F	M00004055D:D05	CH01COH
26	12227	RTA00002909F.e.18.1.P.Seq	F	M00022601B:G06	CH03MAH
27	11047	RTA00002893F.o.06.1.P.Seq	F	M00003960D:C12	CH01COH
28	1870	RTA00002910F.m.08.1.P.Seq	F	M00023020C:H03	CH03MAH
29	20065	RTA00002908F.m.09.1.P.Seq	F	M00022491A:A08	CH03MAH
30	19454	RTA00002900F.m.23.1.P.Seq	F	M00005379A:D10	CH02COH
31	48048	RTA00002922F.m.13.1.P.Seq	F	M00039124D:H01	CH09LNL
32	19799	RTA00002908F.h.19.1.P.Seq	F	M00022449D:F08	CH03MAH
33	185562	RTA00002911F.m.07.1.P.Seq	F	M00027093A:H02	CH04MAL
34	24214	RTA00002891F.k.19.1.P.Seq	F	M00003764D:F07	CH01COH
35	5172	RTA00002908F.p.22.1.P.Seq	F	M00022525B:D09	CH03MAH
36	50495	RTA00002898F.c.16.1.P.Seq	F	M00004321C:C11	CH01COH
37	43287	RTA00002908F.k.16.1.P.Seq	F	M00022470D:B02	CH03MAH
38	15324	RTA00002905F.p.20.1.P.Seq	F	M00021697C:B07	CH03MAH
39	22157	RTA00002888F.g.07.1.P.Seq	F	M00001461D:B10	CH01COH
40	15249	RTA00002915F.l.08.1.P.Seq	F	M00032489B:G12	CH08LNH
41	2764	RTA00002925F.c.11.1.P.Seq	F	M00039829B:E01	CH09LNL
42	23838	RTA00002889F.b.14.1.P.Seq	F	M00001518B:D10	CH01COH
43	11074	RTA00002899F.g.22.1.P.Seq	F	M00004e03C:C10	CH01COH
44	18367	RTA00002922F.b.09.1.P.Seq	F	M00038619D:C12	CH09LNL
45	21703	RTA00002903F.m.08.1.P.Seq	F	M00007059B:D07	CH02COH
46	21470	RTA00002895F.c.14.1.P.Seq	F	M00004067B:D03	CH01COH
47	15492	RTA00002907F.p.06.1.P.Seq	F	M00022282B:C09	CH03MAH
48	4022	RTA00002897F.i.22.1.P.Seq	F	M00004269B:B04	CH01COH
49	21579	RTA00002891F.e.03.1.P.Seq	F	M00001e86B:H01	CH01COH
50	186283	RTA00002913F.c.06.1.P.Seq	F	M00027831B:D07	CH04MAL

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
151	19978	RTA00002908F.k.21.1.P.Seq	F	M00022472D:B01	CH03MAH
152	31133	RTA00002905F.f.23.1.P.Seq	F	M00008045C:A05	CH03MAH
153	4503	RTA00002893F.h.14.1.P.Seq	F	M00003906A:C02	CH01COH
154	9017	RTA00002886F.i.03.1.P.Seq	F	M00001358A:E08	CH01COH
155	6635	RTA00002895F.a.07.1.P.Seq	F	M00004057D:G01	CH01COH
156	10220	RTA00002921F.d.09.1.P.Seq	F	M00033360C:A03	CH09LNL
157	9831	RTA00002896F.b.08.1.P.Seq	F	M00004139B:F01	CH01COH
158	17291	RTA00002888F.c.10.1.P.Seq	F	M00001442A:F08	CH01COH
159	9802	RTA00002916F.l.12.1.P.Seq	F	M00032628C:B06	CH08LNH
160	133828	RTA00002908F.b.21.1.P.Seq	F	M00022374C:E11	CH03MAH
161	25870	RTA00002909F.m.23.1.P.Seq	F	M00022702D:E02	CH03MAH
162	27324	RTA00002912F.i.22.1.P.Seq	F	M00027433B:D12	CH04MAL
163	98159	RTA00002910F.f.19.1.P.Seq	F	M00022897B:F06	CH03MAH
164	21264	RTA00002903F.d.15.1.P.Seq	F	M00006904D:A02	CH02COH
165	186199	RTA00002911F.b.13.2.P.Seq	F	M00023394D:D10	CH04MAL
166	28794	RTA00002887F.i.14.1.P.Seq	F	M00001403C:B03	CH01COH
167	23180	RTA00002895F.j.17.1.P.Seq	F	M00004093A:C03	CH01COH
168	21022	RTA00002922F.l.01.1.P.Seq	F	M00039111A:C12	CH09LNL
169	14370	RTA00002893F.i.17.1.P.Seq	F	M00003911C:A09	CH01COH
170	4804	RTA00002918F.a.19.1.P.Seq	F	M00032826C:D10	CH08LNH
171	7066	RTA00002919F.o.07.1.P.Seq	F	M00033246A:H12	CH08LNH
172	48227	RTA00002903F.o.18.1.P.Seq	F	M00007117A:C11	CH02COH
173	20171	RTA00002886F.i.15.1.P.Seq	F	M00001359A:H10	CH01COH
174	10555	RTA00002894F.p.10.1.P.Seq	F	M00004055C:B10	CH01COH
175	12523	RTA00002914F.m.08.1.P.Seq	F	M00028361B:H08	CH08LNH
176	23767	RTA00002896F.i.21.1.P.Seq	F	M00004171B:B03	CH01COH
177	16849	RTA00002918F.b.07.1.P.Seq	F	M00032829D:A05	CH08LNH
178	185866	RTA00002911F.c.18.2.P.Seq	F	M00026818C:E01	CH04MAL
179	29927	RTA00002899F.b.20.1.P.Seq	F	M00004443C:F07	CH01COH
180	21975	RTA00002902F.a.01.1.P.Seq	F	M00005743D:A12	CH02COH
181	24456	RTA00002903F.b.20.1.P.Seq	F	M00006877C:F11	CH02COH
182	6034	RTA00002901F.a.12.1.P.Seq	F	M00005423A:C11	CH02COH
183	11362	RTA00002887F.h.06.1.P.Seq	F	M00001399C:A01	CH01COH
184	20671	RTA00002905F.a.22.1.P.Seq	F	M00007947A:B06	CH03MAH
185	8059	RTA00002917F.b.02.1.P.Seq	F	M00032671B:D06	CH08LNH
186	12037	RTA00002897F.d.11.1.P.Seq	F	M00004229B:B06	CH01COH
187	13209	RTA00002897F.d.20.1.P.Seq	F	M00004230D:B05	CH01COH
188	23660	RTA00002915F.i.21.1.P.Seq	F	M00031416D:H05	CH08LNH
189	4747	RTA00002919F.c.23.1.P.Seq	F	M00033041A:B11	CH08LNH
190	24532	RTA00002919F.m.16.1.P.Seq	F	M00033218C:F07	CH08LNH
191	8576	RTA00002890F.h.13.1.P.Seq	F	M00001616D:F03	CH01COH
192	12056	RTA00002893F.g.12.1.P.Seq	F	M00003900C:D12	CH01COH
193	895	RTA00002921F.b.11.1.P.Seq	F	M00033303C:F09	CH09LNL
194	7212	RTA00002897F.j.04.1.P.Seq	F	M00004270A:E09	CH01COH
195	108296	RTA00002907F.h.20.1.P.Seq	F	M00022193C:C09	CH03MAH
196	115713	RTA00002906F.a.22.1.P.Seq	F	M00021852C:H02	CH03MAH
197	7334	RTA00002910F.l.08.1.P.Seq	F	M00023004C:A01	CH03MAH
198	1090	RTA00002918F.g.20.2.P.Seq	F	M00032892C:C12	CH08LNH
199	7913	RTA00002886F.j.13.1.P.Seq	F	M00001362A:F09	CH01COH
200	12139	RTA00002923F.o.02.1.P.Seq	F	M00039349D:B11	CH09LNL

SEQ ID	CLUSTER	SEQ NAME	ORIENTATION	CLONE ID	LIBRARY
751	32293	RTA00002901F.i.13.1.P.Seq	F	M00005535B:B01	CH02COH
752	8913	RTA00002901F.j.07.1.P.Seq	F	M00005557D:H10	CH02COH
753	185819	RTA00002912F.a.20.1.P.Seq	F	M00027215A:F06	CH04MAL
754	10559	RTA00002898F.o.12.1.P.Seq	F	M00004406A:G09	CH01COH
755	8740	RTA00002923F.o.11.1.P.Seq	F	M00039383A:H07	CH09LNL
756	160257	RTA00002907F.l.12.2.P.Seq	F	M00022237C:E04	CH03MAH
757	6078	RTA00002930F.c.11.1.P.Seq	F	M00055433D:G03	CH15CON
758	12543	RTA00002927F.b.14.1.P.Seq	F	M00039377B:E05	CH12EDT
759	9686	RTA00002930F.f.19.1.P.Seq	F	M00055794A:E10	CH15CON
760	3369	RTA00002930F.b.12.1.P.Seq	F	M00042732B:H06	CH15CON
761	6891	RTA00002895F.i.03.1.P.Seq	F	M00004087C:E02	CH01COH
762	13666	RTA00002892F.i.05.1.P.Seq	F	M00003822C:A09	CH01COH
763	6925	RTA00002930F.k.24.1.P.Seq	F	M00056458C:E01	CH15CON
764	11351	RTA00002901F.g.15.1.P.Seq	F	M00005504D:F06	CH02COH
765	11497	RTA00002889F.a.21.1.P.Seq	F	M00001512D:F08	CH01COH
766	1596	RTA00002922F.m.18.1.P.Seq	F	M00039125D:H12	CH09LNL
767	186519	RTA00002924F.a.22.1.P.Seq	F	M00039411D:D09	CH09LNL
768	24429	RTA00002903F.j.04.1.P.Seq	F	M00006989B:G05	CH02COH
769	33795	RTA00002902F.k.18.1.P.Seq	F	M00006739B:A04	CH02COH
770	24267	RTA00002889F.l.17.1.P.Seq	F	M00001561D:H04	CH01COH
771	12536	RTA00002891F.j.20.1.P.Seq	F	M00003760C:G10	CH01COH
772	22627	RTA00002887F.k.07.1.P.Seq	F	M00001410A:G10	CH01COH
773	24430	RTA00002901F.h.20.1.P.Seq	F	M00005520B:E01	CH02COH
774	16151	RTA00002897F.l.22.1.P.Seq	F	M00004284A:F08	CH01COH
775	6148	RTA00002890F.i.16.1.P.Seq	F	M00001623D:E12	CH01COH
776	106064	RTA00002908F.l.19.1.P.Seq	F	M00022485B:E07	CH03MAH
777	9573	RTA00002893F.p.13.1.P.Seq	F	M00003970D:H07	CH01COH
778	19542	RTA00002902F.l.20.1.P.Seq	F	M00006756B:G06	CH02COH
779	16672	RTA00002889F.b.21.1.P.Seq	F	M00001528C:C03	CH01COH
780	8573	RTA00002891F.p.07.1.P.Seq	F	M00003785D:F07	CH01COH
781	15746	RTA00002896F.h.10.1.P.Seq	F	M00004163C:A03	CH01COH
782	4500	RTA00002887F.b.08.1.P.Seq	F	M00001387A:C12	CH01COH
783	16003	RTA00002910F.c.08.1.P.Seq	F	M00022820A:F07	CH03MAH
784	18723	RTA00002916F.g.18.1.P.Seq	F	M00032580D:A09	CH08LNH
785	4270	RTA00002922F.b.01.1.P.Seq	F	M00038616C:C09	CH09LNL
786	30095	RTA00002907F.i.20.1.P.Seq	F	M00022208C:E04	CH03MAH
787	42916	RTA00002924F.c.08.1.P.Seq	F	M00039433B:D06	CH09LNL
788	13652	RTA00002902F.j.09.1.P.Seq	F	M00006714C:D06	CH02COH
789	6972	RTA00002902F.j.06.1.P.Seq	F	M00006712C:H01	CH02COH
790	4519	RTA00002910F.i.06.1.P.Seq	F	M00022947B:D02	CH03MAH
791	13106	RTA00002928F.f.09.1.P.Seq	F	M00040224C:F06	CH13EDT
792	98186	RTA00002909F.m.08.1.P.Seq	F	M00022696B:C11	CH03MAH
793	3167	RTA00002898F.g.09.1.P.Seq	F	M00004342D:C12	CH01COH
794	3272	RTA00002897F.a.18.1.P.Seq	F	M00004212D:C03	CH01COH
795	14446	RTA00002899F.d.05.1.P.Seq	F	M00004462D:D12	CH01COH
796	17865	RTA00002918F.a.13.1.P.Seq	F	M00032825B:F08	CH08LNH
797	5834	RTA00002898F.h.12.1.P.Seq	F	M00004352A:D08	CH01COH
798	14533	RTA00002896F.k.24.1.P.Seq	F	M00004179C:B06	CH01COH
799	15222	RTA00002900F.j.05.1.P.Seq	F	M00005332A:C06	CH02COH
800	22594	RTA00002898F.h.21.1.P.Seq	F	M00004357B:B06	CH01COH

Table 3

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
1	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
2	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
3	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
4	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
5	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
6	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
7	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
8	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
9	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
10	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
11	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
12	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
13	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
14	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
15	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
16	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
17	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
18	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
19	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
20	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
21	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
22	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
23	<NONE>	<NONE>	<NONE>	548562	GENOME POLYPROTEIN [CONTAINS: RNA REPLICASE ; HELICASE; COAT PROTEIN] 2.7.7.48) - apple stem grooving virus (strain P-209)	9.2
24	<NONE>	<NONE>	<NONE>	416959	EXCISION REPAIR PROTEIN ERCC-6 DNA repair helicase ERCC6 - human >gi|182181 (L04791) excision repair protein [Homo sapiens]	8.9
25	<NONE>	<NONE>	<NONE>	3327096	(AB014541) KIAA0641 protein [Homo sapiens]	8.7
26	<NONE>	<NONE>	<NONE>	861293	(U28741) F35D2.1 gene product [Caenorhabditis elegans]	7.9
27	<NONE>	<NONE>	<NONE>	3297821	(AL031032) extensin-like protein	5.5
28	<NONE>	<NONE>	<NONE>	2119692	transforming growth factor-beta type III receptor - chicken >gi|511843 (L01121) transforming growth factor-beta type III receptor [Gallus gallus]	5.1
29	<NONE>	<NONE>	<NONE>	2136028	protein-kinase PRK1 - human	5.0

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
173	U72487	Rattus norvegicus calcium-independent alpha-latrotoxin receptor mRNA, complete cds	1.7	544411	GLYCOPROTEIN GP100 PRECURSOR (P29F8) discoideum]	0.35
174	AE000718	Aquifex aeolicus section 50 of 109 of the complete genome	1.7	2497569	FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (FGFR-3) (HEPARIN-BINDING GROWTH FACTOR RECEPTOR) >gi|2117851|pir||I55363 fibroblast growth factor receptor 3 - mouse >gi|199145 (M81342) fibroblast growth factor receptor 3 [Mus musculus]	0.34
175	AF016897	Oryza sativa GDP dissociation inhibitor protein OsGDI2 (OsGDI2) mRNA, complete cds	1.7	125362	MACROPHAGE COLONY STIMULATING FACTOR I RECEPTOR PRECURSOR (CSF-1-R) (FMS PROTO-ONCOGENE) (C-FMS) factor 1 receptor - cat >gi|163855 (J03149) M-CSF receptor [Felis domesticus]	0.34
176	U95102	Xenopus laevis mitotic phosphoprotein 90 mRNA, complete cds	1.7	85058	muscarinic acetylcholine receptor - fruit fly acetylcholine receptor [Drosophila melanogaster]	0.20
177	AF077352	Chlamydomonas reinhardtii myosin heavy chain	1.7	728901	ACROSOMAL PROTEIN SP-10 PRECURSOR SP-10 - western baboon >gi|298488|bbs|127113 (S56458) SP-10=intraacrosomal protein [Papio papio=baboons, Peptide, 285 aa] [Papio hamadryas]	0.20
178	Z92788	Caenorhabditis elegans cosmid F53B8, complete sequence [Caenorhabditis elegans]	1.7	746516	(U23517) D1022.7 [Caenorhabditis elegans] >gi|3258651 elegans]	0.068

SEQ ID	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
775	Y14971	Gallus gallus mRNA for K60 protein	0.022	134091	U1 SMALL NUCLEAR RIBONUCLEOPROTEIN 70 KD (U1 SNRNP 70 KD) >gi|85864|pir||S02016 U1 snRNP 70K protein - African clawed frog >gi|65179 (X12430) U1 70K [Xenopus laevis]	0.032
776	AF003133	Caenorhabditis elegans cosmid T21E3	0.022	1709997	DNA REPAIR PROTEIN RAD18 >gi|1150622 protein rad18 [Schizosaccharomyces pombe]	2e-08
777	AF003133	Caenorhabditis elegans cosmid T21E3	0.022	1709997	DNA REPAIR PROTEIN RAD18 >gi|1150622 protein rad18 [Schizosaccharomyces pombe]	2e-08
778	U57645	Human helix-loop-helix proteins Id-1 (ID-1) and Id-1' (ID-1) genes, complete cds	0.021	<NONE>	<NONE>	<NONE>
779	U67570	Methanococcus jannaschii section 112 of 150 of the complete genome	0.021	<NONE>	<NONE>	<NONE>
780	L01584	Trypanosoma cruzi calcium-binding protein (CUB2.8) gene, complete cds.	0.021	<NONE>	<NONE>	<NONE>
781	L04787	Borrelia hermsii outer membrane lipoprotein	0.021	<NONE>	<NONE>	<NONE>
782	U95094	Xenopus laevis XL-INCENP (XL-INCENP) mRNA, complete cds	0.021	<NONE>	<NONE>	<NONE>
783	L36890	Saccharomyces cerevisiae mitochondrion transfer RNA-Thr1 (tRNA-Thr) gene; transfer RNA-Val (tRNA-Val) gene; oxi2 gene, complete cds; ORF2 and origin of replication (ori5).	0.021	<NONE>	<NONE>	<NONE>

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
1812	AB016930	Cricetulus griseus mRNA for Phosphatidylglycerophosphate synthase, complete cds	6e-96	4159682	(AB016930) Phosphatidylglycerophosphate synthase [Cricetulus griseus]	7e-41
1813	AB005549	Rattus norvegicus mRNA for atypical PKC specific binding protein, complete cds	7e-97	3868778	(AB005549) atypical PKC specific binding protein [Rattus norvegicus]	3e-41
1814	X90849	G.gallus PB1 gene	2e-97	2134381	polybromo 1 protein - chicken chicken >gi|951231 (X90849) polybromo 1 protein [Gallus gallus]	1e-34
1815	S79873	h-lamp-2=lysosome-associated membrane protein-2 protein-2b (LAMP2) mRNA, alternatively spliced form h-lamp-2b, complete cds.	3e-98	<NONE>	<NONE>	<NONE>
1816	U67203	Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds	2e-98	1675224	(U67204) ACF7 neural isoform 2 [Mus musculus]	9e-39
1817	L14684	Rattus norvegicus nuclear-encoded mitochondrial elongation factor G mRNA, complete cds.	e-100	585084	ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G) >gi|543383|pir||S40780 translation elongation factor G, mitochondrial - rat >gi|310102	2e-30
1818	X84692	M.musculus Spnr mRNA for RNA binding protein	e-133	1363238	spermatid perinuclear RNA-binding protein Spnr - mouse >gi|673454 (X84692) spermatid perinuclear RNA binding protein [Mus musculus]	5e-35
1819	U50736	Rattus norvegicus cardiac adriamycin responsive protein mRNA, complete cds	e-113	1362781	cytokine inducible nuclear protein C193 - human >gi|793841 (X83703) nuclear protein [Homo sapiens]	2e-36
1820	S66855	HoxB9=Hox-2.5 [mice, embryos, mRNA Partial, 786 nt]	e-107	1708355	HOMEOBOX PROTEIN HOX-B9 (HOX-2.5)	8e-37

SEQ ID	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
		HoxB9=Hox-2.5				
1821	S66855	[mice, embryos, mRNA Partial, 786 nt]	e-108	1708355	HOMEOBOX PROTEIN HOX-B9 (HOX-2.5)	4e-37
1822	U92072	Rattus norvegicus m-tomosyn mRNA, complete cds	e-102	3790389	(U92072) m-tomosyn [Rattus norvegicus]	2e-38
1823	D17577	Mouse mRNA for kinesin-like protein (Kif1b), complete cds	e-129	2497524	KINESIN-LIKE PROTEIN KIF1B mouse >gi|407339|gnl|PID|d1005029 (D17577) Kif1b [Mus musculus]	2e-39
1824	AF062484	Mus musculus SDP8 mRNA, complete cds	e-122	3126981	(AF062484) SDP8 [Mus musculus]	5e-40
1825	X73683	R.norvegicus mRNA for histone H3.3	e-109	122075	(H3.3Q) histone H3.3 - fruit fly (Drosophila melanogaster) histone H3.3B - chicken >gi|2119023|pir||S61218 histone H3.3 - fruit fly (Drosophila hydei) 1-136) [Oryctolagus cuniculus] >gi|8046 (X53822) Histone H3.3Q gene product [Drosophila melanogaster] >gi|51198 gallus] >gi|161190 (M17876) histone H3 [Spisula solidissima] >gi|211853 (M11393) histone 3.3 [Gallus gallus] >gi|306848 (M11354) H3.3 histone [Homo sapiens] melanogaster] >gi|963031 (X81205) histone H3.3 H3.3A variant [Drosophila melanogaster] musculus]	2e-40
1826	U67203	Mus musculus ACF7 neural isoform 1 (mACF7) mRNA, partial cds	e-102	1675224	(U67204) ACF7 neural isoform 2 [Mus musculus]	2e-40
1827	D17577	Mouse mRNA for kinesin-like protein (Kif1b), complete cds	e-131	2497524	KINESIN-LIKE PROTEIN KIF1B mouse >gi|407339|gnl|PID|d1005029 (D17577) Kif1b [Mus musculus]	7e-42

Table 4

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
1	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
2	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
3	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
4	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
5	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
6	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
7	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
8	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
9	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
10	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
11	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
12	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
13	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
14	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
15	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
16	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
17	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
18	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
19	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
20	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
21	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
22	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>	<NONE>
23	<NONE>	<NONE>	<NONE>	1079469	tMDC I protein - crab-eating macaque	9.3
24	<NONE>	<NONE>	<NONE>	3043656	(AB011138) KIAA0566 protein [Homo sapiens]	9.3
25	<NONE>	<NONE>	<NONE>	112175	potassium channel protein RK5 - rat protein [Rattus norvegicus]	8.6
26	<NONE>	<NONE>	<NONE>	3769624	(AF091565) olfactory receptor [Rattus norvegicus]	7.2
27	<NONE>	<NONE>	<NONE>	3876443	(Z81517) F28B1.6 [Caenorhabditis elegans]	7.1
28	<NONE>	<NONE>	<NONE>	2224464	(AB001684) ORF249 [Chlorella vulgaris]	6.9
29	<NONE>	<NONE>	<NONE>	1519707	(U67940) ORFveg106; random cDNA sequence [Dictyostelium discoideum]	6.7
30	<NONE>	<NONE>	<NONE>	227491	protein kinase C II [Xenopus laevis]	6.7
31	<NONE>	<NONE>	<NONE>	630575	C50C3.4 protein - Caenorhabditis elegans	6.0
32	<NONE>	<NONE>	<NONE>	137290	35 KD PROTEIN IN RNA2 clover necrotic mosaic virus >gi|61466 (X08021) ORF for 35 kDa polypeptide (AA 1-317) [Red clover necrotic mosaic virus]	6.0

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
174	Z46255	S.cerevisiae chromosome VI lambda clone.	1.7	3875228	(Z46792) similar to lethal(1) discs large-1 tumor suppressor protein-like repeats; cDNA EST EMBL:D33495 comes from this gene; cDNA EST EMBL:D35117 comes from this gene; cDNA EST EMBL:D36356 comes from this gene; cDNA EST EMB... >gi|3879984|gnl|PID|e1351767 suppressor protein-like repeats; cDNA EST EMBL:D33495 comes from this gene; cDNA EST EMBL:D35117 comes from this gene; cDNA EST EMBL:D36356 comes from this gene; cDNA EST EMB...	6.7
175	U01066	Human CD4 promoter, partial sequence.	1.7	125448	THYMIDINE KINASE saimiriine herpesvirus 1 (strain 11[Onc]) >gi|60341	6.7
176	U34743	Phalaenopsis sp. 'hybrid SM9108' homeobox protein mRNA, complete cds	1.7	1022918	(U38184) ATPase subunit 6 [Trypanosoma cruzi]	6.7
177	U14662	Baboon herpesvirus HVP2 gB glycoprotein (UL27) gene, complete cds.	1.7	3218378	(AL023862) hypothetical protein SC3F9.07 [Streptomyces coelicolor]	6.7
178	AB017006	Homo sapiens PMS2L15 mRNA, partial cds	1.7	1465855	(U64859) glutamine-rich protein [Caenorhabditis elegans]	6.7
179	U92651	Brassica oleracea var. botrytis tonoplast intrinsic protein bobTIP26-1 mRNA, complete cds	1.7	3023675	DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC) dynein heavy chain [Schizosaccharomyces pombe]	6.6
180	AF000634	Lytechinus variegatus notch homolog mRNA, complete cds	1.7	148574	(M58520) endo-1,4-beta-glucanase [Fibrobacter succinogenes]	6.6

	Nearest Neighbor (BlastN vs. Genbank)			Nearest Neighbor (BlastX vs. Non-Redundant Proteins)		
SEQ ID	ACCESSION	DESCRIPTION	P VALUE	ACCESSION	DESCRIPTION	P VALUE
		Mus musculus inositol polyphosphate 5-phosphatase II (INPP5P) mRNA, complete cds				
772	AF040094		0.022	<NONE>	<NONE>	<NONE>
773	X76776	H.sapiens HLA-DMB gene	0.022	<NONE>	<NONE>	<NONE>
774	AE001521	Helicobacter pylori, strain J99 section 82 of 132 of the complete genome	0.022	<NONE>	<NONE>	<NONE>
775	X16004	A.longa rbcL, rpl5, rps8, rpl36, rps14, rps2, trnI,trnF, trnC and rpoB (partial) genes > :: emb|X75651|ALRIBP A.longa plastid genes for ribosomal proteins, tRNAs, RNA polymerase subunit beta and rubisco large subunit	0.022	<NONE>	<NONE>	<NONE>
776	Y12707	Lactococcus lactis cremoris plasmid pHW393 DNA, rlladii, mlladii genes	0.022	<NONE>	<NONE>	<NONE>
777	U27118	Arabidopsis thaliana glutamyl-tRNA reductase	0.022	<NONE>	<NONE>	<NONE>
778	Z96622	H.sapiens telomeric DNA sequence, clone 5PTEL002, read 5PTELOO002.seq	0.022	191333	(J05503) carbamoyl-phosphate synthetase (E.C.6.3.5.5)	9.8
779	D83984	Sulculus diversicolor DNA for IDO-like myoglobin, complete cds	0.022	1078509	probable membrane protein YDR018c - yeast	9.7
780	Z77952	H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA4A3	0.022	4204206	(AB022786) N-acetyl-beta-D-glucosaminidase [Enterobacter sp.]	7.5

Table 5

SEQ ID	Start	Stop	Score	Direction	Description
29	295	421	5872	For	mkk like kinases
30	31	182	3943	For	Basic region plus leucine zipper transcription factors
31	298	397	5625	For	mkk like kinases
186	175	395	7660	For	SH2 Domain
187	358	432	4320	For	Ank repeat
196	37	322	6049	For	mkk like kinases
234	23	121	4607	For	SH3 Domain
308	110	172	4150	For	Zinc finger, C2H2 type
410	42	191	4036	For	Basic region plus leucine zipper transcription factors
431	71	428	5538	Rev	ATPases Associated with Various Cellular Activities
552	116	288	3930	Rev	Basic region plus leucine zipper transcription factors
639	157	561	5797	For	ATPases Associated with Various Cellular Activities
746	209	427	5379	For	Fibronectin type III domain
768	116	288	3930	For	Basic region plus leucine zipper transcription factors
807	339	392	3620	For	Zinc finger, C2H2 type
820	341	406	2930	Rev	EF-hand
822	108	262	4179	For	Basic region plus leucine zipper transcription factors
836	158	353	4430	For	Basic region plus leucine zipper transcription factors
1157	41	444	5279	Rev	protein kinase
1192	186	416	5469	For	Fibronectin type III domain
1268	238	315	3540	For	Ank repeat
1269	79	240	11640	For	LIM domain containing proteins
1288	73	234	3953	For	Basic region plus leucine zipper transcription factors

SEQ ID	Start	Stop	Score	Direction	Description
1622	180	365	4022	for	Basic region plus leucine zipper transcription factors
1630	100	291	3998	for	Basic region plus leucine zipper transcription factors
1674	196	258	4880	for	Zinc finger, C2H2 type
1676	9	86	6610	for	Homeobox Domain
1677	316	369	5780	rev	Thioredoxins
1688	109	410	17414	for	Ras family
1704	184	372	3977	for	Basic region plus leucine zipper transcription factors
1707	92	439	24100	rev	Phosphatidylinositol-specific phospholipase C, Y domain
1711	263	361	6400	for	WD domain, G-beta repeats
1744	238	433	10572	rev	Serine carboxypeptidases
1755	281	367	2580	for	EF-hand
1762	236	334	5880	for	WD domain, G-beta repeats
1779	64	126	4790	for	Zinc finger, C2H2 type
1801	295	351	4030	for	Zinc finger, C2H2 type
1804	301	378	3460	for	Ank repeat
1808	36	161	4170	for	Basic region plus leucine zipper transcription factors
1811	184	315	8390	for	N-terminal homology in Ets domain
1814	127	294	10770	for	Bromodomain (conserved sequence found in human, Drosophila and yeast proteins.)
1818	9	146	4741	for	Double-stranded RNA binding motif
1819	278	355	3460	for	Ank repeat
1820	123	299	12150	for	Homeobox Domain
1821	127	303	12180	for	Homeobox Domain
1830	184	267	4270	for	Ank repeat
1832	18	173	8987	for	SH3 Domain
1835	51	206	8987	for	SH3 Domain
1839	224	307	4270	for	Ank repeat
1846	12	398	36700	for	G-protein alpha subunit

Example 4

DIFFERENTIAL EXPRESSION OF POLYNUCLEOTIDES OF THE INVENTION: DESCRIPTION OF LIBRARIES AND DETECTION OF DIFFERENTIAL EXPRESSION

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 6 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepare the cDNA library, the abbreviated name of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 6

Description of cDNA Libraries

Library (lib #)	Description	Number of Clones in this Clustering
1	Km12 L4 Human Colon Cell Line, High Metastatic Potential (derived from Km12C) "High Colon"	307133
2	Km12C Human Colon Cell Line, Low Metastatic Potential "Low Colon"	284755
3	MDA-MB-231 Human Breast Cancer Cell Line, High Metastatic Potential; micro-metastases in lung "High Breast"	326937
4	MCF7 Human Breast Cancer Cell, Non Metastatic "Low Breast"	318979
8	MV-522 Human Lung Cancer Cell Line, High Metastatic Potential "High Lung"	223620
9	UCP-3 Human Lung Cancer Cell Line, Low Metastatic Potential "Low Lung"	312503

Library (lib #)	Description	Number of Clones in this Clustering
12	Human microvascular endothelial cells (HMEC) – Untreated PCR (OligodT) cDNA library	41938
13	Human microvascular endothelial cells (HMEC) – Basic fibroblast growth factor (bFGF) treated PCR (OligodT) cDNA library	42100
14	Human microvascular endothelial cells (HMEC) – Vascular endothelial growth factor (VEGF) treated PCR (OligodT) cDNA library	42825
15	Normal Colon – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	34285
16	Colon Tumor – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	35625
17	Liver Metastasis from Colon Tumor of UC#2 Patient PCR (OligodT) cDNA library "High Colon Metastasis Tissue"	36984
18	Normal Colon – UC#3 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	36216
19	Colon Tumor – UC#3 Patient PCR (OligodT) cDNA library "High Colon Tumor Tissue"	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient PCR (OligodT) cDNA library "High Colon Metastasis Tissue"	30956
21	G RRpz Human Prostate Cell Line	164801
22	WOca Human Prostate Cancer Cell Line	162088

The KM12L4 and KM12C cell lines are described in Example 1 above. The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma.

The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran et al., *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar et al., *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson et al., *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang et al., *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki et al., *Int J Cancer* (1987) *40*:46 (UCP-3); Varki et al., *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki et al., *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner et al., *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang et al., *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz cell line refers to low passage (3 passages or fewer) human prostate cells, and the WOca cell line refers to low passage (3 passages or fewer) human prostate cancer cells.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac et al., *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may

occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1st), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2nd). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, "Differences between Proportions," pp 296-298 (1974)).

EXAMPLE 5

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL BREAST CANCER CELLS VERSUS LOW METASTATIC BREAST CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential breast cancer tissue and low metastatic breast cancer cells. Expression of these sequences in breast cancer can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential breast cancer cells and low metastatic potential breast cancer cells.

Table 7

Differentially expressed polynucleotides: Higher expression in high metastatic potential breast cancer (lib3) relative to low metastatic breast cancer cells (lib4)

SEQ ID NOs:	Lib3 clones	Lib4 clones	lib3/lib4
472	64	0	62
1851	6	0	6
1856	8	0	8
1867	6	0	6
1872	6	0	6
1875	12	3	4
1923	89	22	4

SEQ ID NOs:	Lib3 clones	Lib4 clones	lib3/lib4
2118	7	0	7
2119	7	0	7
2135	37	13	3
2190	19	0	19
2193	16	5	3
2232	12	2	6
2239	6	0	6
2338	21	2	10
2378	16	4	4
2394	6	0	6
2395	6	0	6
2490	13	3	4
2505	16	2	8
2540	8	1	8
2542	11	1	11
2607	11	2	5
2640	22	5	4
2674	8	0	8
2679	19	0	19
2684	14	4	3
2707	8	0	8
2724	9	0	9
2757	6	0	6
2776	10	0	10
2804	13	2	6
2818	6	0	6
2906	14	0	14
2959	26	8	3
2964	17	4	4
2968	6	0	6
2977	22	3	7
2980	13	1	13
3010	6	0	6
3043	10	1	10
3071	33	12	3
3072	9	1	9
3095	19	3	6
3097	11	2	5
3173	12	2	6
3203	8	1	8
3210	27	8	3

SEQ ID NOs:	Lib3 clones	Lib4 clones	lib3/lib4
3212	13	1	13
3284	8	0	8
3288	6	0	6
3331	14	3	5
3335	13	1	13

Table 8

Differentially expressed polynucleotides: Higher expression in low metastatic breast cancer cells (lib4) relative to high metastatic potential breast cancer (lib3)

SEQ ID NOs:	Lib 3 Clones	Lib 4 Clones	lib4/lib3
402	0	6	6
614	3	21	7
624	0	6	6
626	0	8	8
712	0	9	9
744	0	7	7
1325	2	29	15
1452	2	13	7
1880	0	9	9
1915	0	7	7
1951	0	6	6
1955	8	32	4
2015	0	7	7
2046	0	7	7
2076	1	22	23
2087	0	6	6
2124	0	9	9
2145	0	8	8
2162	0	6	6
2163	0	12	12
2164	5	19	4
2172	2	15	8
2192	5	16	3
2244	20	43	2
2266	3	18	6
2313	24	56	2
2346	1	13	13

SEQ ID NOs:	Lib 3 Clones	Lib 4 Clones	lib4/lib3
2355	0	10	10
2371	0	6	6
2393	1	17	17
2404	1	21	22
2443	0	6	6
2460	0	11	11
2523	0	6	6
2575	1	10	10
2578	0	6	6
2584	1	17	17
2590	0	6	6
2609	1	9	9
2632	5	24	5
2714	5	24	5
2728	0	6	6
2752	1	14	14
2794	4	15	4
2826	0	7	7
2987	5	15	3
3005	1	14	14
3009	20	58	3
3047	4	17	4
3057	2	17	9
3075	2	11	6
3076	0	6	6
3102	0	6	6
3128	15	52	4
3132	15	52	4
3142	0	6	6
3187	22	49	2
3253	23	96	4
3282	19	46	2
3285	20	40	2
3346	0	9	9

EXAMPLE 6

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL LUNG CANCER CELLS VERSUS LOW METASTATIC LUNG CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential lung cancer cells and low metastatic lung cancer cells. Expression of these sequences in lung cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 9

Differentially expressed polynucleotides: Higher expression in high metastatic potential lung cancer cells (lib8) relative to low metastatic lung cancer cells (lib9)

SEQ ID NO:	Lib8 clones	Lib9 clones	lib8/lib9
14	10	0	10
137	5	0	5
151	5	0	7
152	9	0	13
171	6	0	8
200	10	0	14
254	5	0	7
262	5	0	7
271	5	0	7
348	6	1	8
412	5	0	7
507	5	0	7
520	6	0	8
530	5	0	7
588	5	0	7
623	7	0	10
637	7	0	10
660	5	0	7
678	8	0	11
680	5	0	7
700	9	2	6
714	28	13	3
774	11	0	15
812	5	0	7
834	8	2	6
901	11	2	8
1168	5	0	7
1333	6	0	8
1352	5	0	7
1524	11	1	15
1706	5	0	7
1752	17	9	3
1768	20	4	7
1769	5	0	7
1780	6	0	8

SEQ ID NO:	Lib8 clones	Lib9 clones	lib8/lib9
1781	40	3	19
1799	6	1	8
1803	6	1	8
1811	16	9	2
1884	6	0	8
1919	8	1	11
1939	6	0	8
1975	43	9	7
2024	12	1	17
2045	8	1	11
2060	20	13	2
2071	16	4	6
2128	5	0	7
2177	10	2	7
2181	44	13	5
2184	11	1	15
2185	10	4	3
2283	7	0	10
2311	10	4	3
2314	10	0	14
2393	14	6	3
2398	6	1	8
2460	10	4	3
2514	6	0	8
2597	5	0	7
2657	8	2	6
2669	6	1	8
2670	6	1	8
3047	21	3	10
3050	16	5	4
3092	7	1	10
3140	181	119	2
3157	5	0	7
3187	16	5	4
3210	5	0	7
3220	28	4	10
3236	7	1	10
3249	16	0	22
3264	8	2	6
3305	7	0	10
3309	20	0	28

SEQ ID NO:	Lib8 clones	Lib9 clones	lib8/lib9
3318	24	4	8
3330	5	0	7
3331	5	0	7

Table 10

Differentially expressed polynucleotides: Higher expression in low metastatic lung cancer cells (lib 9) relative to high metastatic potential lung cancer cells (lib 8)

SEQ ID NO:	Lib 8 clones	Lib 9 clones	lib 9/lib 8
24	3	20	5
53	0	18	13
64	0	8	6
70	0	11	8
105	10	66	5
129	0	16	11
214	1	14	10
233	4	35	6
237	0	13	9
264	0	29	21
329	2	17	6
368	1	37	26
370	0	11	8
418	0	8	6
450	0	9	6
461	0	9	6
484	0	26	19
494	0	41	29
517	1	12	9
522	1	11	8
581	1	17	12
614	3	23	5
706	0	11	8
726	5	23	3
806	0	14	10
824	0	9	6
836	1	14	10
874	0	12	9
900	5	21	3
1017	2	14	5

SEQ ID NO:	Lib 8 clones	Lib 9 clones	lib 9/lib 8
1144	0	8	6
1154	0	12	9
1166	2	45	16
1170	1	13	9
1302	2	13	5
1326	1	13	9
1327	1	13	9
1367	0	12	9
1377	0	12	9
1437	2	18	6
1442	1	14	10
1466	0	13	9
1476	0	13	9
1495	0	8	6
1496	1	13	9
1664	38	253	5
1682	1	17	12
1687	0	9	6
1758	0	8	6
1817	4	18	3
1837	3	16	4
1845	3	23	5
1856	2	17	6
1910	1	18	13
2146	2	16	9
2156	0	9	6
2463	0	12	9
2724	10	38	3
2749	403	2000	4
2801	6	25	3
2993	3	18	4
3080	0	10	7
3107	3	23	5
3292	0	20	14
3324	110	548	4

EXAMPLE 7

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL COLON CANCER CELLS VERSUS LOW METASTATIC COLON CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer cells and low metastatic colon cancer cells. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and low metastatic potential colon cancer cells:

Table 11

Differentially expressed polynucleotides: Higher expression in low metastatic colon cancer cells (lib 2) relative to high metastatic potential colon cancer cells (lib 1)

SEQ ID NOs:	Lib 1 clones	Lib 2 clones	lib 2/lib 1
429	0	9	10
1494	0	8	9
1923	34	114	4
1986	3	12	4
2018	0	9	10
2036	2	10	5
2049	8	25	3
2135	24	87	4

SEQ ID NOs:	Lib 1 clones	Lib 2 clones	lib 2/lib 1
2146	2	16	9
2208	6	27	5
2215	2	11	6
2239	1	10	11
2307	2	12	6
2313	28	62	2
2357	5	14	3
2360	3	21	8
2362	0	6	6
2378	3	12	4
2569	3	20	7
2571	0	6	6
2588	54	172	3
2592	15	41	3
2611	0	6	6
2636	0	9	10
2641	7	20	3
2650	0	9	10
2662	0	9	10
2674	4	13	4
2682	0	6	6
2702	9	25	3
2704	8	23	3
2715	2	12	6
2804	9	22	3
2821	13	29	2
2840	1	8	9
2846	2	15	8
2866	0	6	6
2906	0	6	6
2915	44	109	3
2933	0	6	6
2935	5	16	3
2957	1	11	12
2959	3	27	10
2977	16	30	2
2980	12	27	2
3000	2	13	7
3009	12	29	3
3115	0	7	8
3156	502	2170	5

SEQ ID NOs:	Lib 1 clones	Lib 2 clones	lib 2/lib 1
3210	2	21	11
3211	0	9	10
3213	0	7	8
3235	2	12	6
3251	2	12	6
3296	3	12	4
3335	1	8	9

EXAMPLE 8

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH METASTATIC POTENTIAL COLON CANCER PATIENT TISSUE VERSUS NORMAL PATIENT TISSUE

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the advanced disease state which involves processes such as angiogenesis, dedifferentiation, cell replication, and metastasis. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer tissue and normal colon tissue:

Table 12

Differentially expressed polynucleotides isolated from samples from two patients (patient 2 and patient 3 and) : Lower expression in high metastatic potential colon tissue (patient 2:lib 17; patient 3:lib 20) vs. normal colon tissue (patient 2:lib 15; patient 3:lib 18)

SEQ ID NO:	lib 15 clones	lib 17 clones	lib 15/lib 17
69	19	7	3
123	6	0	6
140	24	8	3
197	6	0	6
198	113	0	121
254	28	9	3
412	28	9	3
512	11	1	12
641	17	7	3
642	7	0	8
954	12	3	4
1011	209	16	14
1024	8	0	9
1040	12	3	4
1055	26	7	4
1106	31	15	2
1125	17	0	18
1129	17	0	18
1138	109	0	117
1244	14	1	15
1253	73	0	78
1283	34	7	5
1285	34	7	5
1339	13	4	3
1474	73	0	78
1505	18	3	6
1553	68	6	12
1554	2542	14	195
1605	2542	14	195
1628	6	0	6
1643	142	4	38
1753	12	0	10
1764	13	0	14

SEQ ID NO:	lib 15 clones	lib 17 clones	lib 15/lib 17
SEQ ID NO:	Lib18 Clones	Lib20 Clones	lib18/lib20
105	28	11	2
198	21	0	18
254	9	0	8
412	9	0	8
1011	11	1	9
1138	14	0	12
1253	23	0	20
1643	18	0	15
1764	12	0	10
3156	140	43	3

Table 13

Differentially expressed polynucleotides isolated from samples from two patients (patient 2 and patient 3): Lower expression in normal colon tissue (patient 2:lib 15; patient 3:lib 18)vs. high metastatic potential colon tissue (patient 2:lib 17; patient 3:lib 20).

SEQ ID NO:	Lib 15 Clones	Lib 17 Clones	lib 17/lib 15
321	3	23	7
363	1	9	8
836	21	99	4
859	6	20	3
885	13	28	2
916	13	28	2
981	2	11	5
1226	8	70	8
1308	0	8	7
1317	29	84	3
1429	27	127	4
1442	0	9	8
1534	1	12	11
1540	12	43	3
1552	0	7	7
1556	1	9	8
1557	1	9	8
1569	2189	5122	2
1571	6	18	3
1576	3	25	8

SEQ ID NO:	Lib 15 Clones	Lib 17 Clones	lib 17/lib 15
1581	4	22	5
1601	25	157	6
1613	9	48	5
1616	15	61	4
1620	2	17	8
1622	4	99	23
1626	6	35	5
1647	4	22	5
1664	4	28	7
1683	2	18	8
1704	3	15	5
1800	0	7	7
2749	23	60	2
2784	4	14	3
2805	1	9	8
2976	3	14	4
3128	18	57	3
3129	26	124	4
3146	64	210	3
3150	940	2267	2
3151	2	15	7
SEQ ID NO:	lib 18 clones	lib 20 clones	lib 20/lib 18
865	0	5	6
1569	1	7	8
1580	1	7	8
1590	1	7	8
2790	0	5	6

EXAMPLE 9

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH COLON TUMOR POTENTIAL PATIENT TISSUE VERSUS METASTASIZED COLON CANCER PATIENT TISSUE

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from colon cancer tissue and cells derived from colon cancer tissue metastases to liver. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information associated with the transformation of precancerous tissue to malignant tissue. This information

can be useful in the prevention of achieving the advanced malignant state in these tissues, and can be important in risk assessment for a patient.

The following table summarizes identified polynucleotides with differential expression between high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells:

Table 14

Differentially expressed polynucleotides:
Greater expression in metastatic colon tumor tissue (lib 20) vs.
colon tumor tissue (lib 19)

SEQ ID NO:	lib 19 clones	lib 20 clones	lib 20/lib 19
937	0	6	8
976	0	5	7
1520	1	8	11
1546	1	11	15
1550	1	11	15
1574	1	8	11
1580	0	7	9
1590	0	7	9
1599	8	21	4
1607	158	632	5
1622	1	7	9

Table 15

Greater expression in colon tumor tissue (lib 19) than metastatic colon tissue (lib 20)

SEQ ID NO:	lib 19 clones	lib 20 clones	lib 19/lib 20
105	64	11	4
1011	53	1	40
1226	18	4	3
1571	8	0	6
1726	15	3	4
1811	17	2	6
2749	47	6	6
3146	19	2	7
3324	20	1	15

EXAMPLE 10

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN HIGH TUMOR POTENTIAL COLON CANCER PATIENT TISSUE VERSUS NORMAL PATIENT TISSUE

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information associated with the prevention of the malignant state in these tissues, and can be important in risk assessment for a patient. For example, sequences that are highly expressed in the potential colon cancer cells are associated with or can be indicative of increased expression of genes or regulatory sequences involved in early tumor progression. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer cells and normal colon cells:

Table 16

Differentially expressed polynucleotides detected in samples from patient (patient 2)
Higher expression in normal colon tissue (patient 2, lib 15)
vs. tumor potential colon tissue (patient 2:lib16)

SEQ ID NO:	lib 15 clones	lib 16 clones	lib 16/lib 15
69	19	7	3
105	116	54	2
140	24	4	6
197	6	0	6
198	113	3	40
254	28	6	5
412	28	6	5
642	7	0	7
830	10	2	5
938	31	13	3
1011	209	37	6
1095	12	3	4
1125	17	0	18

SEQ ID NO:	lib 15 clones	lib 16 clones	lib 16/lib 15
1129	17	0	18
1138	109	1	115
1253	73	1	77
1283	34	13	3
1285	34	13	3
1339	13	3	5
1453	11	3	4
1474	73	1	77
1505	18	6	3
1554	2542	448	6
1605	2542	448	6
1614	36	14	3
1630	24	9	3
1643	142	2	75
1646	39	14	3
1649	24	8	3
1677	19	6	3
1753	13	0	14
1764	13	0	14
1766	177	65	3
1772	24	8	3

Table 17

Differentially expressed polypeptides detected in samples from patient. Lower expression in normal colon tissue (lib 18) than colon tumor tissue (lib 19)

SEQ ID NO:	lib 18 clones	lib 19 clones	lib 19/lib 18
3146	3	19	6
3150	21	228	10
3324	3	20	6

5

Table 18

Differentially expressed polypeptides detected in samples from patient. Higher expression in normal colon tissue (lib 18) than colon tumor tissue (lib 19)

SEQ ID NO:	lib 18 clones	lib 19 clones	lib 18/lib 19
198	21	2	12
465	6	0	7
489	6	0	7
745	6	0	7
859	11	2	6
976	7	0	8
1011	209	37	6
1045	8	1	9
1138	14	0	16
1253	23	0	26
1392	16	4	5
1474	23	0	26
1589	6	0	7
1591	22	11	2
1607	386	158	3
1643	18	0	21
1753	12	0	14
1764	12	0	14
SEQ ID NO:	lib 18 clones	lib 19 clones	lib 19/lib 18
105	28	64	2
1011	11	53	4
1226	2	18	8
1251	6	19	3
1559	1	9	8
1571	0	8	7
1608	1	9	8
1766	2	13	6
1782	1	9	8
1811	1	17	15

Table 19

Differentially expressed polynucleotides:

Higher expression in colon tumor tissue

(patient 2, lib 16) vs. normal colon tissue (patient 2, lib 15)

SEQ ID NO:	lib 15 clones	lib 16 clones	lib 16/lib 15
7	1	9	9
164	6	19	3
734	4	15	4
836	21	53	2
928	2	11	5
965	2	11	5
987	2	11	5
1026	7	19	3
1044	4	16	4
1119	4	16	4
1226	8	46	5
1227	0	9	9
1251	7	95	13
1316	0	6	6
1429	27	81	3
1442	0	9	9
1540	12	28	2
1553	68	590	8
1560	4	24	6
1577	1	10	9
1588	5	20	4
1610	3	13	4
1620	2	23	11
1626	6	23	4
1673	2	15	7
2416	0	7	7
2749	23	54	2
2976	3	14	4
3129	26	64	2
3132	18	54	3

5

EXAMPLE 11

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN GROWTH FACTOR-STIMULATED HUMAN MICROVASCULAR ENDOTHELIAL CELLS (HMEC) RELATIVE TO UNTREATED HMEC

A number of polynucleotide sequences have been identified that are differentially expressed between human microvascular endothelial cells (HMEC) that have been treated with growth factors relative to untreated HMEC.

Sequences that are differentially expressed between growth factor-treated HMEC and untreated HMEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other developmental and oncogenic processes. For example, sequences that are more highly expressed in HMEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMEC can serve as markers of cancer cells of higher metastatic potential. Detection of expression of these sequences in colon cancer tissue can provide diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following table summarizes identified polynucleotides with differential expression between growth factor-treated and untreated HMEC.

Table 20

Differentially expressed polynucleotides:

Higher expression in untreated HMEC (lib 12) vs. bFGF treated HMEC (lib 13)

SEQ ID NO:	lib 12 clones	lib 13 clones	lib 12/lib 13
849	6	0	6
1059	6	0	6
1206	12	2	6
3208	12	0	12

Lower expression in untreated HMEC (lib 12) vs. bFGF treated HMEC (lib 13)

2748	3	12	4
3325	0	6	6

Table 21

Differentially expressed polynucleotides:

Higher expression in untreated HMEC (lib 12) VEGF treated HMEC (lib14)

SEQ ID NO:	lib 12 clones	lib 14 clones	lib 12/lib 14
1150	9	0	9

Lower expression in untreated HMEC (lib 12) vs. VEGF treated HMEC (lib14)

3324	22	50	2

EXAMPLE 12

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED IN NORMAL PROSTATE CELLS RELATIVE TO PROSTATE CANCER CELLS

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from normal prostate cells and prostate cancer cells. Expression of these sequences prostate tissue suspected of being cancerous can provide diagnostic, prognostic and/or treatment information. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers. The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and low metastatic potential colon cancer cells:

Table 22

Differentially expressed polynucleotides: normal prostate cell line (lib 21) vs. prostate cancer cell line (lib 22)

Higher in lib 21

SEQ ID NO:	lib 21 clones	lib 22 clones	lib 21/lib 22
53	17	2	8
1754	22	8	3
1801	7	0	7
1845	22	6	4
446	8	0	8
1410	6	0	6
2060	18	6	3
2143	12	3	4
2632	13	1	13
2899	16	2	8
3338	12	2	6

Higher in lib 22

86	2	13	7
93	0	9	9
687	0	9	9
1269	1	15	15
1581	25	74	3
1647	25	74	3
1649	12	27	2
1710	5	16	3
1717	5	16	3
1772	12	27	2
1960	0	6	6
2987	0	6	6
3128	13	42	3
3132	13	42	3
3150	263	962	4
3222	0	6	6
3268	0	6	6

EXAMPLE 13

POLYNUCLEOTIDES DIFFERENTIALLY EXPRESSED ACROSS MULTIPLE LIBRARIES

A number of polynucleotide sequences have been identified that are differentially expressed between cancerous cells and normal cells across two or more tissue types tested (*i.e.*, breast, colon, lung, and prostate). Expression of these sequences in a tissue of any origin can provide diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The following polynucleotides were differentially expressed but without tissue type-specificity in at least two of the breast, colon, lung, and prostate libraries tested: 53, 105, 355, 412, 614, 836, 1442, 1581, 1647, 1649, 1664, 1772, 1782, 1811, 1845, 1856, 1875, 1923, 2060, 2071, 2135, 2146, 2239, 2313, 2378, 2393, 2416, 2460, 2490, 2632, 2674, 2704, 2724, 2749, 2784, 2804, 2959, 2976, 2977, 2980, 2987, 3009, 3047, 3128, 3129, 3132, 3146, 3150, 3156, 3210, 3324, 3331, and 3335.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information:

The following materials were deposited with the American Type Culture Collection (ATCC); CMCC = Chiron Master Culture Collection:

cDNA Libraries Deposited with ATCC

Tube Number	Deposit Date	ATCC Accession No.	CMCC Accession No.
ES137	May 30, 2000		
ES138	May 30, 2000		
ES139	May 30, 2000		
ES140	May 30, 2000		
ES141	May 30, 2000		
ES142	May 30, 2000		
ES143	May 30, 2000		
ES144	May 30, 2000		
ES145	May 30, 2000		
ES146	May 30, 2000		
ES147	May 30, 2000		
ES148	May 30, 2000		
ES149	May 30, 2000		
ES150	May 30, 2000		
ES151	May 30, 2000		
ES152	May 30, 2000		
ES153	May 30, 2000		
ES154	May 30, 2000		
ES155	May 30, 2000		
ES156	May 30, 2000		
ES157	May 30, 2000		
ES158	May 30, 2000		
ES159	May 30, 2000		
ES160	May 30, 2000		
ES161	May 30, 2000		
ES162	May 30, 2000		
ES163	May 30, 2000		
ES164	May 30, 2000		
ES165	May 30, 2000		
ES166	May 30, 2000		
ES167	May 30, 2000		

Table 23 lists the clones for each deposit, designated as "tube" number. This deposit is provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences

herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Retrieval of Individual Clones from Deposit of Pooled Clones

Where the ATCC deposit is composed of a pool of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Table 23

Clone Name	Tube
M00001351A:B02	ES 137
M00001356A:H11	ES 137
M00001363D:D09	ES 137
M00001395D:H02	ES 137
M00001439C:H06	ES 137
M00001476B:G10	ES 137
M00001582A:E02	ES 137
M00003750D:E06	ES 137
M00003761C:F02	ES 137
M00003770A:E05	ES 137
M00003786A:A11	ES 137
M00003800A:F09	ES 137
M00003816D:E11	ES 137
M00003902A:C03	ES 137
M00003991C:F06	ES 137

Clone Name	Tube
M00003995B:E03	ES 137
M00004046C:A08	ES 137
M00004105D:D05	ES 137
M00004139B:B10	ES 137
M00004140D:C03	ES 137
M00004144A:H05	ES 137
M00004152A:C12	ES 137
M00004155D:A10	ES 137
M00004168A:G11	ES 137
M00004197B:H10	ES 137
M00004222C:E03	ES 137
M00004234A:E07	ES 137
M00004239B:F11	ES 137
M00004241B:H07	ES 137
M00004264B:A05	ES 137

Clone Name	Tube
M00004278A:F09	ES 137
M00004282D:C11	ES 137
M00004308C:C06	ES 137
M00004340C:C07	ES 137
M00004354D:E05	ES 137
M00004361A:H02	ES 137
M00004372B:F07	ES 137
M00004378A:B10	ES 137
M00004393B:E07	ES 137
M00023282A:C02	ES 137
M00023300D:C11	ES 137
M00023316C:G08	ES 137
M00023333D:C12	ES 137
M00023352B:F03	ES 137
M00023352D:H03	ES 137
M00023376B:G04	ES 137
M00023377B:F01	ES 137
M00023398B:D12	ES 137
M00023399C:E10	ES 137
M00026803A:F08	ES 137
M00026843B:D10	ES 137
M00026850D:F09	ES 137
M00026851B:F01	ES 137
M00026856D:F02	ES 137
M00026857D:G12	ES 137
M00026859D:D01	ES 137
M00026860B:C05	ES 137
M00026865B:A06	ES 137
M00026868C:E11	ES 137
M00026878A:F05	ES 137
M00026882D:G09	ES 137
M00026885A:H09	ES 137
M00026901A:G07	ES 137
M00026914A:H10	ES 137
M00026915B:C06	ES 137
M00026918B:D01	ES 137
M00026922C:B02	ES 137
M00026922C:G03	ES 137
M00026926A:E10	ES 137
M00026927D:F02	ES 137
M00026928D:A03	ES 137
M00026935C:B04	ES 137
M00026941D:A04	ES 137
M00026944B:E03	ES 137
M00026946A:F12	ES 137

Clone Name	Tube
M00026980A:D09	ES 137
M00027016A:B06	ES 137
M00027018A:C09	ES 137
M00027021A:G02	ES 137
M00027022D:G11	ES 137
M00027030C:H06	ES 137
M00027035D:C06	ES 137
M00027049B:F05	ES 137
M00027078A:B02	ES 137
M00027080A:B01	ES 137
M00027085C:E11	ES 137
M00027094A:B03	ES 137
M00027103B:A09	ES 137
M00027108C:B03	ES 137
M00027121D:C05	ES 137
M00027135A:B11	ES 137
M00027136C:C09	ES 137
M00027141C:H03	ES 137
M00027159D:F03	ES 137
M00027162B:F05	ES 137
M00027178B:G09	ES 137
M00027179D:E06	ES 138
M00027181D:A05	ES 138
M00027195C:E04	ES 138
M00027198B:B08	ES 138
M00027200A:F02	ES 138
M00027207B:F07	ES 138
M00027212D:E03	ES 138
M00027228D:A01	ES 138
M00027232D:B08	ES 138
M00027233B:C01	ES 138
M00027236A:E04	ES 138
M00027237C:B08	ES 138
M00027248A:C02	ES 138
M00027256B:H09	ES 138
M00027258A:A07	ES 138
M00027263A:F10	ES 138
M00027292D:F10	ES 138
M00027297A:C04	ES 138
M00027299B:B12	ES 138
M00027301A:G05	ES 138
M00027301B:B08	ES 138
M00027314C:D09	ES 138
M00027319D:B11	ES 138
M00027324D:C05	ES 138

Clone Name	Tube
M00027347C:G07	ES 138
M00027355A:B07	ES 138
M00027359B:G05	ES 138
M00027366A:F11	ES 138
M00027379C:B07	ES 138
M00027392B:H02	ES 138
M00027396D:G08	ES 138
M00027398C:F07	ES 138
M00027438C:G07	ES 138
M00027462A:D07	ES 138
M00027462B:H07	ES 138
M00027468A:C09	ES 138
M00027475B:E10	ES 138
M00027476A:C09	ES 138
M00027486A:F06	ES 138
M00027520A:C05	ES 138
M00027525B:D06	ES 138
M00027526D:F03	ES 138
M00027528C:B10	ES 138
M00027537C:B01	ES 138
M00027546C:B10	ES 138
M00027591B:C04	ES 138
M00027596A:A10	ES 138
M00027596C:E06	ES 138
M00027602B:C01	ES 138
M00027615A:F10	ES 138
M00027617B:C12	ES 138
M00027620D:F11	ES 138
M00027625A:H01	ES 138
M00027634A:D11	ES 138
M00027641C:A03	ES 138
M00027647C:D03	ES 138
M00027652B:F11	ES 138
M00027668C:H12	ES 138
M00027729D:H06	ES 138
M00027733A:A02	ES 138
M00027741B:F09	ES 138
M00027743A:C03	ES 138
M00027801C:C11	ES 138
M00027813C:F01	ES 138
M00027818C:C07	ES 138
M00027836D:F12	ES 138
M00027837C:D09	ES 138
M00028120D:F12	ES 138
M00028066C:D07	ES 138

Clone Name	Tube
M00028184D:G10	ES 138
M00028185B:A06	ES 138
M00028196D:A03	ES 138
M00028201B:H12	ES 138
M00028207D:E09	ES 138
M00028210B:D02	ES 138
M00028212C:B08	ES 138
M00028215D:F03	ES 138
M00028220A:B04	ES 138
M00028314D:F05	ES 138
M00028316B:H12	ES 138
M00028354A:B12	ES 138
M00028354D:A03	ES 138
M00028357A:G10	ES 138
M00028362A:G11	ES 138
M00028364C:G08	ES 138
M00028369D:E08	ES 138
M00028617C:A12	ES 138
M00028768C:D05	ES 138
M00028770A:D04	ES 138
M00028772C:B09	ES 138
M00028775D:F03	ES 138
M00028777B:G12	ES 138
M00031368A:E10	ES 138
M00031417C:G09	ES 138
M00031419D:C04	ES 138
M00031485D:G02	ES 138
M00032480B:E10	ES 139
M00032492A:C01	ES 139
M00032495B:D02	ES 139
M00032499C:A01	ES 139
M00032508B:H03	ES 139
M00032510D:F12	ES 139
M00032510D:G06	ES 139
M00032513D:F01	ES 139
M00032530D:C02	ES 139
M00032535D:H01	ES 139
M00032539B:C11	ES 139
M00032540A:A09	ES 139
M00032541D:H08	ES 139
M00032545B:H09	ES 139
M00032545D:G05	ES 139
M00032550D:C02	ES 139
M00032551B:G05	ES 139
M00032577A:C04	ES 139

Clone Name	Tube
M00032578A:G06	ES 139
M00032584A:H08	ES 139
M00032592A:H11	ES 139
M00032597C:B01	ES 139
M00032638C:G08	ES 139
M00032638D:A06	ES 139
M00032668D:G12	ES 139
M00032678C:D06	ES 139
M00032688D:D11	ES 139
M00032712B:G02	ES 139
M00032724A:C05	ES 139
M00032725C:F06	ES 139
M00032726C:C01	ES 139
M00032731B:C10	ES 139
M00032731C:C07	ES 139
M00032737B:E09	ES 139
M00032739A:A06	ES 139
M00032744B:F10	ES 139
M00032766B:D12	ES 139
M00032766C:A04	ES 139
M00032790B:A07	ES 139
M00032793A:F06	ES 139
M00032797B:G02	ES 139
M00032808B:G10	ES 139
M00032811B:D02	ES 139
M00032829B:E06	ES 139
M00032830D:G03	ES 139
M00032831C:G07	ES 139
M00032853D:G12	ES 139
M00032864B:B09	ES 139
M00032871D:E11	ES 139
M00032876C:D06	ES 139
M00032907A:G04	ES 139
M00032909A:B06	ES 139
M00032917D:G09	ES 139
M00032918B:D08	ES 139
M00032918B:E06	ES 139
M00032918C:B10	ES 139
M00032921B:H08	ES 139
M00032933A:C10	ES 139
M00032939B:E07	ES 139
M00032940A:C02	ES 139
M00032942D:C12	ES 139
M00032944B:B02	ES 139
M00032984C:G05	ES 139

Clone Name	Tube
M00032990B:A11	ES 139
M00032994A:A08	ES 139
M00032995C:C05	ES 139
M00033007C:E01	ES 139
M00033019B:E10	ES 139
M00033033C:H01	ES 139
M00033034C:A06	ES 139
M00033034C:F02	ES 139
M00033037D:C11	ES 139
M00033074A:C08	ES 139
M00033130B:F06	ES 139
M00033140D:F06	ES 139
M00033173D:C01	ES 139
M00033176B:E12	ES 139
M00033186C:D11	ES 139
M00033189D:F08	ES 139
M00033202D:G06	ES 139
M00033204B:A07	ES 139
M00033205A:F03	ES 139
M00033217B:H07	ES 139
M00033218A:C04	ES 139
M00033223B:H07	ES 139
M00033226A:A11	ES 139
M00033231D:B09	ES 139
M00033231D:G10	ES 139
M00033243B:A05	ES 139
M00033246C:E08	ES 139
M00033248A:B02	ES 139
M00033261C:D12	ES 139
M00033262D:A11	ES 139
M00033263B:G04	ES 139
M00033276B:G08	ES 139
M00033185C:D01	ES 139
M00033288B:D12	ES 140
M00033300D:H12	ES 140
M00033306D:G08	ES 140
M00033306D:H09	ES 140
M00033308B:G05	ES 140
M00033343C:H08	ES 140
M00033345D:A09	ES 140
M00033346C:A05	ES 140
M00033347C:F02	ES 140
M00033349D:F05	ES 140
M00033358A:H12	ES 140
M00033362C:C05	ES 140

Clone Name	Tube
M00033375A:G04	ES 140
M00033376A:C12	ES 140
M00033377D:A05	ES 140
M00033410B:C09	ES 140
M00033424B:A04	ES 140
M00033424D:H12	ES 140
M00033425A:C10	ES 140
M00033427D:F01	ES 140
M00033432B:H10	ES 140
M00033437C:A07	ES 140
M00033437C:C03	ES 140
M00033442A:D06	ES 140
M00033446C:G08	ES 140
M00033446D:B02	ES 140
M00033450C:A02	ES 140
M00033451A:H01	ES 140
M00033454A:D09	ES 140
M00033457D:A05	ES 140
M00033560D:G07	ES 140
M00033561C:A02	ES 140
M00033566C:E08	ES 140
M00033570B:C08	ES 140
M00033570B:E06	ES 140
M00033570C:C10	ES 140
M00033578D:G02	ES 140
M00033581C:H10	ES 140
M00033581D:D08	ES 140
M00033583B:E06	ES 140
M00033583D:B05	ES 140
M00033584D:G11	ES 140
M00033585D:A02	ES 140
M00033588C:G04	ES 140
M00033594C:B03	ES 140
M00033595A:C11	ES 140
M00038259A:G08	ES 140
M00038259B:A02	ES 140
M00038259B:G08	ES 140
M00038259C:H09	ES 140
M00038272A:G01	ES 140
M00038272D:F11	ES 140
M00038279C:A11	ES 140
M00038284B:H04	ES 140
M00038303A:C03	ES 140
M00038303C:D02	ES 140
M00038303D:E07	ES 140

Clone Name	Tube
M00038315C:G11	ES 140
M00038325D:F12	ES 140
M00038326B:D04	ES 140
M00038327A:C11	ES 140
M00038327D:A05	ES 140
M00038328D:A03	ES 140
M00038329A:E08	ES 140
M00038387B:A07	ES 140
M00038614C:H11	ES 140
M00038615A:H12	ES 140
M00038616D:B12	ES 140
M00038618C:C08	ES 140
M00038619B:A03	ES 140
M00038620B:E09	ES 140
M00038631C:B10	ES 140
M00038631D:B02	ES 140
M00038632C:B09	ES 140
M00038633A:D07	ES 140
M00038633B:G02	ES 140
M00038635A:G09	ES 140
M00038635B:C08	ES 140
M00038638D:H03	ES 140
M00038639B:C03	ES 140
M00038639D:F07	ES 140
M00038661A:A07	ES 140
M00038662B:A12	ES 140
M00038663B:H06	ES 140
M00038663D:H10	ES 140
M00038664C:E04	ES 140
M00038991A:D01	ES 140
M00038994A:A10	ES 140
M00038995C:G08	ES 140
M00038995D:E05	ES 140
M00038999B:G11	ES 140
M00038999D:C11	ES 140
M00039002D:G11	ES 140
M00039004B:A06	ES 140
M00039004B:C11	ES 140
M00039005C:H01	ES 141
M00039006D:B01	ES 141
M00039011D:C10	ES 141
M00039013A:C09	ES 141
M00039013D:F02	ES 141
M00039014A:H10	ES 141
M00039014B:C04	ES 141

Clone Name	Tube
M00039015A:D07	ES 141
M00039015B:G10	ES 141
M00039015B:H09	ES 141
M00039015D:H04	ES 141
M00039016A:A02	ES 141
M00039016D:G06	ES 141
M00039024B:B10	ES 141
M00039025A:H09	ES 141
M00039026D:F05	ES 141
M00039028C:B11	ES 141
M00039030B:E02	ES 141
M00039036C:B05	ES 141
M00039039B:E03	ES 141
M00039039B:F09	ES 141
M00039042B:B02	ES 141
M00039043B:E01	ES 141
M00039049D:G07	ES 141
M00039050A:H10	ES 141
M00039052C:F07	ES 141
M00039058A:A04	ES 141
M00039058C:H02	ES 141
M00039059C:G08	ES 141
M00039061B:F08	ES 141
M00039063B:D08	ES 141
M00039064D:H09	ES 141
M00039066D:G08	ES 141
M00039068B:B04	ES 141
M00039068C:E06	ES 141
M00039070D:C02	ES 141
M00039072C:C03	ES 141
M00039072C:E02	ES 141
M00039079A:A05	ES 141
M00039080C:H06	ES 141
M00039081B:G06	ES 141
M00039082B:A05	ES 141
M00039084C:G07	ES 141
M00039084C:H03	ES 141
M00039084C:H04	ES 141
M00039084D:D07	ES 141
M00039096A:A05	ES 141
M00039096A:E07	ES 141
M00039097D:D06	ES 141
M00039099A:H08	ES 141
M00039104D:C09	ES 141
M00039105C:B08	ES 141

Clone Name	Tube
M00039107C:E04	ES 141
M00039108D:B06	ES 141
M00039112B:C05	ES 141
M00039118B:C05	ES 141
M00039118D:A06	ES 141
M00039120C:C09	ES 141
M00039120C:H03	ES 141
M00039123A:B10	ES 141
M00039124C:F03	ES 141
M00039124C:H02	ES 141
M00039124C:H08	ES 141
M00039126D:A08	ES 141
M00039127A:G11	ES 141
M00039127D:E10	ES 141
M00039129C:D04	ES 141
M00039133B:F08	ES 141
M00039135D:F05	ES 141
M00039135D:G02	ES 141
M00039135D:H02	ES 141
M00039139A:C09	ES 141
M00039139C:G12	ES 141
M00039140A:B08	ES 141
M00039140D:A04	ES 141
M00039140D:D09	ES 141
M00039141C:E01	ES 141
M00039142D:B11	ES 141
M00039144C:E06	ES 141
M00039147A:F10	ES 141
M00039156A:B11	ES 141
M00039158B:G12	ES 141
M00039166B:G06	ES 141
M00039167B:H09	ES 141
M00039168C:A04	ES 141
M00039169A:E12	ES 141
M00039170A:B10	ES 141
M00039170C:F05	ES 141
M00039171B:D11	ES 141
M00039177B:D03	ES 141
M00039179A:G09	ES 141
M00039180A:A07	ES 141
M00039196B:H06	ES 141
M00039196D:A07	ES 141
M00039200A:C10	ES 141
M00039211A:C12	ES 141
M00039212C:C12	ES 142

Clone Name	Tube
M00039213A:D01	ES 142
M00039213B:F05	ES 142
M00039218A:F03	ES 142
M00039221A:H03	ES 142
M00039224A:E12	ES 142
M00039228A:B05	ES 142
M00039230A:A10	ES 142
M00039230D:D09	ES 142
M00039230D:G12	ES 142
M00039233A:A03	ES 142
M00039238A:B12	ES 142
M00039238D:A08	ES 142
M00039241A:E11	ES 142
M00039249A:C12	ES 142
M00039249C:G11	ES 142
M00039255C:E12	ES 142
M00039257D:C03	ES 142
M00039258B:E06	ES 142
M00039258D:B08	ES 142
M00039260C:G03	ES 142
M00039263D:A12	ES 142
M00039266A:B02	ES 142
M00039266D:F12	ES 142
M00039266D:H04	ES 142
M00039273B:F02	ES 142
M00039273D:B02	ES 142
M00039274B:G07	ES 142
M00039276B:H09	ES 142
M00039277D:G10	ES 142
M00039279B:C11	ES 142
M00039279B:H02	ES 142
M00039279C:B08	ES 142
M00039281D:B04	ES 142
M00039284D:B12	ES 142
M00039286A:C06	ES 142
M00039287C:A06	ES 142
M00039288C:B11	ES 142
M00039293A:H04	ES 142
M00039293B:C11	ES 142
M00039295B:D03	ES 142
M00039297C:H08	ES 142
M00039298B:B06	ES 142
M00039298B:D03	ES 142
M00039298D:B04	ES 142
M00039299B:G12	ES 142

Clone Name	Tube
M00039300C:C09	ES 142
M00039300C:G04	ES 142
M00039301B:F06	ES 142
M00039303C:F11	ES 142
M00039304D:B09	ES 142
M00039308B:G08	ES 142
M00039310A:C07	ES 142
M00039313D:G04	ES 142
M00039316A:C01	ES 142
M00039318B:B09	ES 142
M00039319B:H12	ES 142
M00039319C:A04	ES 142
M00039322A:F04	ES 142
M00039328D:D07	ES 142
M00039329A:C01	ES 142
M00039329C:B10	ES 142
M00039333D:D09	ES 142
M00039334B:E03	ES 142
M00039335A:E08	ES 142
M00039339A:H07	ES 142
M00039339C:F03	ES 142
M00039340A:D05	ES 142
M00039340B:E07	ES 142
M00039340B:G08	ES 142
M00039341C:H11	ES 142
M00039341D:D07	ES 142
M00039343B:F12	ES 142
M00039344B:G07	ES 142
M00039345A:D09	ES 142
M00039345C:C12	ES 142
M00039381C:H08	ES 142
M00039381D:C02	ES 142
M00039384C:E02	ES 142
M00039384C:F08	ES 142
M00039385B:E09	ES 142
M00039391D:F08	ES 142
M00039396D:B04	ES 142
M00039397B:H09	ES 142
M00039398A:B10	ES 142
M00039401B:D02	ES 142
M00039402B:E03	ES 142
M00039403A:G12	ES 142
M00039404B:A05	ES 142
M00039407B:G02	ES 142
M00039411C:E07	ES 142

Clone Name	Tube
M00039412D:G06	ES 142
M00039414D:G03	ES 142
M00039415D:E01	ES 142
M00039417A:D03	ES 142
M00039417A:E12	ES 142
M00039417B:F01	ES 143
M00039417C:A01	ES 143
M00039417C:G01	ES 143
M00039418B:D08	ES 143
M00039420D:D03	ES 143
M00039422D:F04	ES 143
M00039425C:G01	ES 143
M00039425D:E12	ES 143
M00039428C:E01	ES 143
M00039430B:F12	ES 143
M00039431B:F04	ES 143
M00039432C:A01	ES 143
M00039444C:H02	ES 143
M00039452C:G09	ES 143
M00039454B:A11	ES 143
M00039455D:H04	ES 143
M00039456A:C08	ES 143
M00039458B:H11	ES 143
M00039461A:F04	ES 143
M00039465A:A08	ES 143
M00039472C:B08	ES 143
M00039475C:E10	ES 143
M00039476B:A02	ES 143
M00039477A:B03	ES 143
M00039477D:A10	ES 143
M00039611D:D11	ES 143
M00039612B:B10	ES 143
M00039612B:G05	ES 143
M00039616A:B10	ES 143
M00039616B:C01	ES 143
M00039619B:D02	ES 143
M00039631A:C10	ES 143
M00039633D:D05	ES 143
M00039636C:D11	ES 143
M00039637C:A10	ES 143
M00039652B:D05	ES 143
M00039655B:H09	ES 143
M00039655C:C07	ES 143
M00039655C:E08	ES 143
M00039660C:C10	ES 143

Clone Name	Tube
M00039663C:G09	ES 143
M00039664D:G07	ES 143
M00039672D:D10	ES 143
M00039673A:F09	ES 143
M00039675D:B03	ES 143
M00039675D:H05	ES 143
M00039677A:B08	ES 143
M00039681B:H09	ES 143
M00039682A:C08	ES 143
M00039682C:H11	ES 143
M00039684D:B08	ES 143
M00039685A:A08	ES 143
M00039686C:C05	ES 143
M00039686C:E06	ES 143
M00039688C:G06	ES 143
M00039689C:E08	ES 143
M00039696A:E05	ES 143
M00039697B:F11	ES 143
M00039700B:D02	ES 143
M00039702A:B12	ES 143
M00039702A:B02	ES 143
M00039705D:F02	ES 143
M00039707A:D02	ES 143
M00039710C:G03	ES 143
M00039720D:D02	ES 143
M00039727C:B09	ES 143
M00039729A:A10	ES 143
M00039771C:E11	ES 143
M00039773D:A09	ES 143
M00039773D:F11	ES 143
M00039774C:A03	ES 143
M00039774C:C09	ES 143
M00039775A:A09	ES 143
M00039777C:E05	ES 143
M00039778B:G03	ES 143
M00039778C:A04	ES 143
M00039781D:D10	ES 143
M00039782A:H10	ES 143
M00039785D:G05	ES 143
M00039788A:E03	ES 143
M00039788B:A06	ES 143
M00039788C:A01	ES 143
M00039790B:D03	ES 143
M00039792A:B04	ES 143
M00039793D:C05	ES 143

Clone Name	Tube
M00039794A:E04	ES 143
M00039795B:H10	ES 143
M00039795D:E10	ES 143
M00039795D:G06	ES 143
M00039797C:G05	ES 143
M00039798B:B02	ES 143
M00039799A:D10	ES 143
M00039801A:H11	ES 143
M00039807A:D01	ES 143
M00039808D:H02	ES 143
M00039810A:H10	ES 143
M00039813B:B01	ES 144
M00039813B:D11	ES 144
M00039815C:F09	ES 144
M00039816B:D04	ES 144
M00039816C:D05	ES 144
M00039820A:F11	ES 144
M00039820A:H11	ES 144
M00039820B:B06	ES 144
M00039827B:F07	ES 144
M00039828B:C05	ES 144
M00039832A:B12	ES 144
M00039835A:F07	ES 144
M00039838A:F05	ES 144
M00039839B:B01	ES 144
M00039839C:E05	ES 144
M00039847A:F06	ES 144
M00039851B:G11	ES 144
M00039851C:D12	ES 144
M00039854B:F09	ES 144
M00039855C:F01	ES 144
M00039857B:G10	ES 144
M00039859A:F06	ES 144
M00039859C:G10	ES 144
M00039864A:A07	ES 144
M00039866B:A08	ES 144
M00039869B:F06	ES 144
M00039875D:A10	ES 144
M00039876D:H09	ES 144
M00039877C:C03	ES 144
M00039879C:F05	ES 144
M00039879D:B11	ES 144
M00039880A:H11	ES 144
M00039884A:H11	ES 144
M00039885C:D01	ES 144

Clone Name	Tube
M00039887C:E07	ES 144
M00039887D:C04	ES 144
M00039888B:D03	ES 144
M00039890A:H05	ES 144
M00039894C:H07	ES 144
M00039896C:H01	ES 144
M00039897D:C10	ES 144
M00039898A:A08	ES 144
M00039898D:C06	ES 144
M00039903A:H07	ES 144
M00039903C:D01	ES 144
M00039903C:F03	ES 144
M00039909C:G05	ES 144
M00039909D:C02	ES 144
M00039910C:G10	ES 144
M00039914D:G12	ES 144
M00039915D:C11	ES 144
M00039927A:F04	ES 144
M00039928B:G05	ES 144
M00039936C:C05	ES 144
M00039938C:A08	ES 144
M00039938C:E11	ES 144
M00039940A:D07	ES 144
M00039940D:G08	ES 144
M00039973D:C08	ES 144
M00039973D:D12	ES 144
M00039975C:C11	ES 144
M00039976D:A12	ES 144
M00039978A:G03	ES 144
M00039981A:E08	ES 144
M00039982C:H04	ES 144
M00039983D:A06	ES 144
M00039984A:C02	ES 144
M00039984B:G12	ES 144
M00039984D:G12	ES 144
M00039987A:F09	ES 144
M00039987C:E12	ES 144
M00039987C:G08	ES 144
M00039988A:E06	ES 144
M00039990C:D10	ES 144
M00040004D:B03	ES 144
M00040005B:C11	ES 144
M00040005D:B07	ES 144
M00040007D:A06	ES 144
M00040009D:B07	ES 144

Clone Name	Tube
M00040010A:F10	ES 144
M00040014B:D01	ES 144
M00040014D:D10	ES 144
M00040014D:F03	ES 144
M00040015C:F08	ES 144
M00040016C:H12	ES 144
M00040017A:C06	ES 144
M00040017D:G03	ES 144
M00040019A:E01	ES 144
M00040021A:F09	ES 144
M00040022C:D06	ES 144
M00040026B:F06	ES 144
M00040029A:B03	ES 144
M00040029A:G04	ES 144
M00040031A:E06	ES 144
M00040032A:B03	ES 144
M00040032A:D09	ES 144
M00040037A:E11	ES 145
M00040038D:G04	ES 145
M00040039D:D06	ES 145
M00040040A:A06	ES 145
M00040041C:C09	ES 145
M00040042B:A10	ES 145
M00040047C:F05	ES 145
M00040052D:F12	ES 145
M00040055D:A06	ES 145
M00040055D:B01	ES 145
M00040060C:H10	ES 145
M00040062B:B05	ES 145
M00040070B:B07	ES 145
M00040071B:A10	ES 145
M00040072C:G09	ES 145
M00040076C:D06	ES 145
M00040077D:C11	ES 145
M00040080C:C06	ES 145
M00040081C:E01	ES 145
M00040085D:A10	ES 145
M00040085D:E04	ES 145
M00040087D:F08	ES 145
M00040088C:E10	ES 145
M00040089A:G08	ES 145
M00040089B:E04	ES 145
M00040089C:E06	ES 145
M00040090B:G09	ES 145
M00040092B:F05	ES 145

Clone Name	Tube
M00040093B:C02	ES 145
M00040093D:D03	ES 145
M00040097A:C12	ES 145
M00040098C:B01	ES 145
M00040098D:E04	ES 145
M00040098D:G12	ES 145
M00040100C:E05	ES 145
M00040100D:B06	ES 145
M00040103B:H10	ES 145
M00040105C:F11	ES 145
M00040106B:B09	ES 145
M00040107B:H07	ES 145
M00040111C:D05	ES 145
M00040115B:A04	ES 145
M00040115B:H12	ES 145
M00040118D:G10	ES 145
M00040121B:C05	ES 145
M00040122D:A02	ES 145
M00040123A:A09	ES 145
M00040124D:H01	ES 145
M00040129D:E10	ES 145
M00040302C:A04	ES 145
M00040304B:F06	ES 145
M00040305A:D11	ES 145
M00040305C:H06	ES 145
M00040307B:F01	ES 145
M00040307C:F10	ES 145
M00040309A:E11	ES 145
M00004825D:D05	ES 145
M00004832D:H02	ES 145
M00004839C:H02	ES 145
M00005018A:B05	ES 145
M00005297D:H08	ES 145
M00005308A:D06	ES 145
M00005351C:G05	ES 145
M00005352C:A02	ES 145
M00005358B:B06	ES 145
M00005359A:D04	ES 145
M00005379A:E04	ES 145
M00005382B:F08	ES 145
M00005384A:C11	ES 145
M00005402B:F08	ES 145
M00005445D:B01	ES 145
M00005449B:B10	ES 145
M00005449B:D01	ES 145

Clone Name	Tube
M00005457C:A03	ES 145
M00005458A:F11	ES 145
M00005498A:H06	ES 145
M00005531D:F06	ES 145
M00005539D:G01	ES 145
M00005555A:A10	ES 145
M00005556B:D02	ES 145
M00005601D:D08	ES 145
M00005614B:B01	ES 145
M00005623A:G02	ES 145
M00005623D:G12	ES 145
M00005625A:C02	ES 145
M00005673B:B12	ES 145
M00005778B:F09	ES 145
M00005805D:D12	ES 145
M00005820C:E04	ES 145
M00006581D:F08	ES 145
M00006599D:B02	ES 145
M00006657C:G05	ES 145
M00006680B:D02	ES 145
M00006712C:H09	ES 145
M00006809B:B09	ES 145
M00006861B:F09	ES 145
M00006866A:D07	ES 146
M00006886D:H02	ES 146
M00006893C:E07	ES 146
M00006897A:H02	ES 146
M00006928D:D07	ES 146
M00006935C:F06	ES 146
M00006968A:G08	ES 146
M00006977C:G04	ES 146
M00006977D:A03	ES 146
M00007012D:H08	ES 146
M00007013A:D09	ES 146
M00007026B:H09	ES 146
M00007108B:A02	ES 146
M00007112C:B10	ES 146
M00007116C:G02	ES 146
M00007124D:H10	ES 146
M00007136A:A03	ES 146
M00007149A:G02	ES 146
M00007157C:F11	ES 146
M00007165B:G11	ES 146
M00007194A:B09	ES 146
M00007929C:B08	ES 146

Clone Name	Tube
M00007941D:C09	ES 146
M00007943D:C09	ES 146
M00007972B:H12	ES 146
M00007976A:C10	ES 146
M00007992C:F06	ES 146
M00007994A:G02	ES 146
M00008006B:B03	ES 146
M00008026B:C11	ES 146
M00008045A:H02	ES 146
M00008053A:F10	ES 146
M00008063B:A06	ES 146
M00021665B:F12	ES 146
M00021671D:F12	ES 146
M00021852D:A05	ES 146
M00021866D:A03	ES 146
M00021908D:G12	ES 146
M00021919C:A10	ES 146
M00021923C:D11	ES 146
M00021955A:H02	ES 146
M00021964C:E10	ES 146
M00021972D:C11	ES 146
M00022005C:C06	ES 146
M00022015B:B07	ES 146
M00022054A:H03	ES 146
M00022084D:B01	ES 146
M00022099B:D06	ES 146
M00022105C:C12	ES 146
M00022127C:H03	ES 146
M00022135C:B05	ES 146
M00022138A:E05	ES 146
M00022175D:D12	ES 146
M00022178B:D06	ES 146
M00022181C:D01	ES 146
M00022183B:C02	ES 146
M00022184C:C11	ES 146
M00022233C:A12	ES 146
M00022234C:D06	ES 146
M00022247A:E02	ES 146
M00022257A:B09	ES 146
M00022262D:G03	ES 146
M00022264B:G10	ES 146
M00022363C:G12	ES 146
M00022365D:A03	ES 146
M00022373A:B05	ES 146
M00022373C:B07	ES 146

Clone Name	Tube
M00022391B:E01	ES 146
M00022391D:F10	ES 146
M00022416A:A07	ES 146
M00022421B:C11	ES 146
M00022433A:E02	ES 146
M00022434D:D06	ES 146
M00022440B:E01	ES 146
M00022444D:G01	ES 146
M00022467C:B12	ES 146
M00022489C:G04	ES 146
M00022492C:A02	ES 146
M00022495D:H08	ES 146
M00022496B:E12	ES 146
M00022499A:B02	ES 146
M00022533A:A08	ES 146
M00022579C:C11	ES 146
M00022597D:A06	ES 146
M00022602A:E09	ES 146
M00022615D:G05	ES 146
M00022634D:C08	ES 146
M00022640C:C12	ES 146
M00022641C:H05	ES 146
M00022646A:H10	ES 146
M00022662D:G11	ES 146
M00022667D:B02	ES 146
M00022668B:B12	ES 146
M00022670D:H11	ES 146
M00022671B:A08	ES 146
M00022684A:C02	ES 146
M00022731A:D02	ES 147
M00022739A:B03	ES 147
M00022747D:E03	ES 147
M00022767B:G11	ES 147
M00022785C:G06	ES 147
M00022793D:B01	ES 147
M00022795B:G06	ES 147
M00022797B:G08	ES 147
M00022817A:H02	ES 147
M00022821C:C09	ES 147
M00022823C:C01	ES 147
M00022830D:D01	ES 147
M00022834B:G11	ES 147
M00022854A:B03	ES 147
M00022856C:A07	ES 147
M00022860C:G04	ES 147

Clone Name	Tube
M00022885C:H05	ES 147
M00022895A:H08	ES 147
M00022910A:A06	ES 147
M00022925C:A08	ES 147
M00022928B:C01	ES 147
M00022930C:E02	ES 147
M00022938B:F07	ES 147
M00022968B:E02	ES 147
M00022976C:F04	ES 147
M00022979A:D05	ES 147
M00022986D:H09	ES 147
M00022997A:F06	ES 147
M00023001C:C08	ES 147
M00023003C:D07	ES 147
M00023007A:H04	ES 147
M00023007C:E10	ES 147
M00023020C:G08	ES 147
M00023024D:F12	ES 147
M00023032A:B05	ES 147
M00023039D:B05	ES 147
M00023042D:D02	ES 147
M00023044B:D02	ES 147
M00023094A:B11	ES 147
M00023100A:E12	ES 147
M00039181D:E05	ES 147
M00039184A:D03	ES 147
M00039184B:B09	ES 147
M00039361B:E01	ES 147
M00039363A:C09	ES 147
M00039366C:B07	ES 147
M00039367B:H02	ES 147
M00039371B:H06	ES 147
M00039372C:D12	ES 147
M00039374B:B07	ES 147
M00039374C:H12	ES 147
M00039374C:H02	ES 147
M00039376D:H07	ES 147
M00039377D:E12	ES 147
M00039378D:H07	ES 147
M00039379A:B03	ES 147
M00039380C:C09	ES 147
M00039482B:G02	ES 147
M00039493A:C04	ES 147
M00039496B:D08	ES 147
M00039496B:H09	ES 147

Clone Name	Tube
M00039497C:C06	ES 147
M00039499C:A04	ES 147
M00039500C:C04	ES 147
M00039505C:E03	ES 147
M00039508A:C12	ES 147
M00039508C:G01	ES 147
M00039510C:G02	ES 147
M00039512C:D06	ES 147
M00039515A:A06	ES 147
M00039515D:C11	ES 147
M00039517B:G12	ES 147
M00039521A:A02	ES 147
M00039521D:H03	ES 147
M00039528B:B12	ES 147
M00039529C:D07	ES 147
M00039530B:E02	ES 147
M00039533A:C12	ES 147
M00039533B:G08	ES 147
M00039533D:F04	ES 147
M00039535D:D10	ES 147
M00039536C:C10	ES 147
M00039536C:H11	ES 147
M00039561A:B07	ES 147
M00039561B:A09	ES 147
M00039562B:G02	ES 147
M00039564B:C01	ES 147
M00039570A:D10	ES 147
M00039570B:D10	ES 147
M00039584C:C01	ES 147
M00039584C:C11	ES 147
M00039587C:F12	ES 147
M00039590D:D02	ES 147
M00039591C:D06	ES 147
M00039595C:E05	ES 147
M00039597D:F04	ES 147
M00039600A:A11	ES 148
M00039604B:E05	ES 148
M00039604D:G03	ES 148
M00039606B:D08	ES 148
M00039607D:E08	ES 148
M00039608D:H01	ES 148
M00039609D:F07	ES 148
M00039624A:H09	ES 148
M00039624B:F12	ES 148
M00039625B:G08	ES 148

Clone Name	Tube
M00039626D:F04	ES 148
M00039629B:F01	ES 148
M00039629D:B04	ES 148
M00039630A:C08	ES 148
M00039630C:H04	ES 148
M00039641A:A05	ES 148
M00039641C:D07	ES 148
M00039642D:B12	ES 148
M00039642D:H09	ES 148
M00039643C:B04	ES 148
M00039645C:E01	ES 148
M00039647A:H11	ES 148
M00039736D:G08	ES 148
M00039740B:F10	ES 148
M00039752B:G08	ES 148
M00039755A:B08	ES 148
M00039756B:H06	ES 148
M00039760B:B08	ES 148
M00040131B:D11	ES 148
M00040131C:F03	ES 148
M00040131D:G08	ES 148
M00040133B:B03	ES 148
M00040136C:F08	ES 148
M00040138B:H03	ES 148
M00040141D:F05	ES 148
M00040143A:H05	ES 148
M00040145D:D03	ES 148
M00040147D:H11	ES 148
M00040160B:A10	ES 148
M00040162A:E01	ES 148
M00040169B:F08	ES 148
M00040173D:B05	ES 148
M00040174C:E10	ES 148
M00040174D:G03	ES 148
M00040181B:H09	ES 148
M00040181D:H10	ES 148
M00040182D:D06	ES 148
M00040183A:F07	ES 148
M00040184C:A11	ES 148
M00040191A:B09	ES 148
M00040221A:G11	ES 148
M00040222D:G02	ES 148
M00040223A:C05	ES 148
M00040226A:H10	ES 148
M00040230A:H02	ES 148

Clone Name	Tube
M00040231B:C08	ES 148
M00040232D:B07	ES 148
M00040233A:H02	ES 148
M00040233C:G05	ES 148
M00040252C:C06	ES 148
M00040253C:A05	ES 148
M00040254B:C10	ES 148
M00040256A:A06	ES 148
M00040257D:H10	ES 148
M00040260B:D02	ES 148
M00040260C:D04	ES 148
M00040261C:F01	ES 148
M00040262B:B06	ES 148
M00040264D:G05	ES 148
M00040265D:B07	ES 148
M00040265D:C08	ES 148
M00040267A:E06	ES 148
M00040267C:C04	ES 148
M00040271B:E12	ES 148
M00040271C:D08	ES 148
M00040273B:H12	ES 148
M00040274A:D07	ES 148
M00040274A:H11	ES 148
M00040280C:H05	ES 148
M00040281D:B01	ES 148
M00040282A:A03	ES 148
M00040286C:C02	ES 148
M00040287C:B09	ES 148
M00040287D:D07	ES 148
M00039746C:A08	ES 148
M00039746C:G09	ES 148
M00039746C:H05	ES 148
M00039746C:H06	ES 148
M00039746D:D11	ES 148
M00039748A:F11	ES 148
M00039748C:F11	ES 148
M00039749D:D05	ES 148
M00039761D:E10	ES 148
M00039762B:F07	ES 148
M00039764C:D07	ES 148
M00039766A:G07	ES 148
M00039766D:H01	ES 149
M00039767B:A04	ES 149
M00039767C:E12	ES 149
M00039770A:G11	ES 149

Clone Name	Tube
M00039770C:E04	ES 149
M00039942D:C01	ES 149
M00039943B:F10	ES 149
M00039945C:F09	ES 149
M00039946B:F08	ES 149
M00039947A:D06	ES 149
M00039947C:G03	ES 149
M00039948A:E03	ES 149
M00039948D:D11	ES 149
M00039951A:B07	ES 149
M00039951B:B12	ES 149
M00039951B:C03	ES 149
M00039955C:C04	ES 149
M00039957C:C09	ES 149
M00039957D:A12	ES 149
M00039958A:A08	ES 149
M00039958C:B09	ES 149
M00040201C:G11	ES 149
M00040202A:F05	ES 149
M00040203A:H06	ES 149
M00040203B:A05	ES 149
M00040203D:H11	ES 149
M00040206A:A07	ES 149
M00040207B:D08	ES 149
M00040208A:C03	ES 149
M00040208B:A07	ES 149
M00040208D:G09	ES 149
M00040217D:B07	ES 149
M00040218C:C02	ES 149
M00040219B:D02	ES 149
M00040219D:E08	ES 149
M00040291D:C05	ES 149
M00040293D:G04	ES 149
M00040294D:D12	ES 149
M00040296D:E09	ES 149
M00040298B:G02	ES 149
M00040299B:F10	ES 149
M00040313C:D05	ES 149
M00040313D:E04	ES 149
M00040314D:H05	ES 149
M00040317A:H03	ES 149
M00040317D:F02	ES 149
M00040318A:B02	ES 149
M00040318C:H11	ES 149
M00040320D:F02	ES 149

Clone Name	Tube
M00040323B:C12	ES 149
M00040323C:G11	ES 149
M00040326A:F04	ES 149
M00040327B:G06	ES 149
M00040332D:B05	ES 149
M00040333D:G05	ES 149
M00040334D:B02	ES 149
M00040334D:C07	ES 149
M00040342B:D12	ES 149
M00040345D:A09	ES 149
M00040346A:C11	ES 149
M00040347D:F09	ES 149
M00040349D:B09	ES 149
M00040351B:F02	ES 149
M00040351D:A11	ES 149
M00040364A:E05	ES 149
M00040366A:B01	ES 149
M00040368A:A12	ES 149
M00040368A:F01	ES 149
M00040368D:E09	ES 149
M00040371C:H05	ES 149
M00040375C:B06	ES 149
M00040376C:G02	ES 149
M00040377C:G07	ES 149
M00040383A:H02	ES 149
M00040383D:C04	ES 149
M00040385C:D02	ES 149
M00040386A:A02	ES 149
M00040387C:E07	ES 149
M00040387D:H05	ES 149
M00040390A:H02	ES 149
M00040390B:F02	ES 149
M00040391A:D10	ES 149
M00040392B:H01	ES 149
M00040392C:B12	ES 149
M00040394A:D04	ES 149
M00040395B:D11	ES 149
M00042534A:A05	ES 149
M00042538B:E06	ES 149
M00042543C:G04	ES 149
M00042558A:F03	ES 149
M00042560A:F12	ES 149
M00042565C:A08	ES 149
M00042566C:C05	ES 149
M00042567B:H10	ES 149

Clone Name	Tube
M00042693D:E04	ES 149
M00042696B:E05	ES 149
M00042697D:C07	ES 150
M00042698D:D10	ES 150
M00042698D:E01	ES 150
M00042702B:G02	ES 150
M00042704A:F09	ES 150
M00042711B:A11	ES 150
M00042717A:C07	ES 150
M00042737C:H04	ES 150
M00042740A:E09	ES 150
M00042742D:D05	ES 150
M00042887C:D07	ES 150
M00042895A:D10	ES 150
M00042895C:G01	ES 150
M00042902D:B08	ES 150
M00042904B:E07	ES 150
M00042905A:F11	ES 150
M00042905B:C03	ES 150
M00042905D:D02	ES 150
M00042347D:H11	ES 150
M00042348B:E05	ES 150
M00042349D:D07	ES 150
M00042431B:G08	ES 150
M00042431C:F01	ES 150
M00042431D:C10	ES 150
M00042432D:E02	ES 150
M00042435A:A11	ES 150
M00042436B:H09	ES 150
M00042437A:D04	ES 150
M00042439B:B03	ES 150
M00042439B:D03	ES 150
M00042440B:E09	ES 150
M00042463A:F09	ES 150
M00042470C:E05	ES 150
M00042511A:H04	ES 150
M00042515C:F08	ES 150
M00042751C:C12	ES 150
M00042752A:E11	ES 150
M00042756B:F11	ES 150
M00042756D:A10	ES 150
M00042759B:G11	ES 150
M00042760A:C12	ES 150
M00042765C:D04	ES 150
M00042767B:G10	ES 150

Clone Name	Tube
M00042769C:E09	ES 150
M00042770B:B12	ES 150
M00042770C:C04	ES 150
M00042771C:F06	ES 150
M00042774C:C05	ES 150
M00042781A:A07	ES 150
M00042784A:H06	ES 150
M00042788C:F11	ES 150
M00042790C:C07	ES 150
M00042792A:H01	ES 150
M00042797D:D10	ES 150
M00042799D:F08	ES 150
M00042800A:A03	ES 150
M00042802C:C04	ES 150
M00042806C:F07	ES 150
M00042807D:D05	ES 150
M00042823C:C02	ES 150
M00042830B:E02	ES 150
M00042839B:B11	ES 150
M00042841D:H07	ES 150
M00042849D:F11	ES 150
M00042852B:A03	ES 150
M00042852C:A01	ES 150
M00042856B:H02	ES 150
M00042352C:H03	ES 150
M00042352D:C01	ES 150
M00042352D:G09	ES 150
M00042448A:C09	ES 150
M00042448C:H12	ES 150
M00042453B:G09	ES 150
M00042518D:A06	ES 150
M00042518D:D04	ES 150
M00043296B:G09	ES 150
M00043304B:D05	ES 150
M00043304C:D02	ES 150
M00043305B:G02	ES 150
M00043306C:B03	ES 150
M00043306D:B07	ES 150
M00043310C:G06	ES 150
M00043311C:E03	ES 150
M00043312C:E08	ES 150
M00043320B:A07	ES 150
M00043324D:H11	ES 150
M00043328D:H02	ES 150
M00043332C:G04	ES 150

Clone Name	Tube
M00043334B:A10	ES 150
M00043338B:A03	ES 150
M00043338B:C11	ES 150
M00043339A:F11	ES 150
M00043340B:H08	ES 150
M00043344D:E04	ES 150
M00043345C:A06	ES 150
M00043346A:G01	ES 150
M00043350D:B11	ES 151
M00043351D:A11	ES 151
M00043352D:B05	ES 151
M00043352D:C03	ES 151
M00043359B:D10	ES 151
M00043359C:G01	ES 151
M00043361B:A01	ES 151
M00043366A:A02	ES 151
M00043366C:H05	ES 151
M00043367B:A08	ES 151
M00043368C:F09	ES 151
M00043370B:C08	ES 151
M00043372C:G05	ES 151
M00043377A:C03	ES 151
M00043378A:H10	ES 151
M00043379D:H02	ES 151
M00043383C:F12	ES 151
M00043383D:A02	ES 151
M00043384B:B02	ES 151
M00043386A:B08	ES 151
M00043389C:E03	ES 151
M00043389D:D07	ES 151
M00043391A:C10	ES 151
M00043391A:G08	ES 151
M00043392D:C11	ES 151
M00043393A:B08	ES 151
M00043401D:G08	ES 151
M00043402C:D08	ES 151
M00043405A:D11	ES 151
M00043405C:G12	ES 151
M00043405C:G02	ES 151
M00043406B:G12	ES 151
M00043407C:E05	ES 151
M00043408B:D11	ES 151
M00043409B:B03	ES 151
M00043410C:A09	ES 151
M00043411B:D08	ES 151

Clone Name	Tube
M00043411D:H06	ES 151
M00042584B:C10	ES 151
M00042623D:D07	ES 151
M00042625C:B04	ES 151
M00042626B:D08	ES 151
M00042627C:D01	ES 151
M00042630A:C05	ES 151
M00042955C:D05	ES 151
M00042956C:B06	ES 151
M00042960D:H08	ES 151
M00042962D:C05	ES 151
M00042964D:A03	ES 151
M00042966B:F07	ES 151
M00042966C:E06	ES 151
M00042970C:A04	ES 151
M00042970C:H10	ES 151
M00042976A:H04	ES 151
M00042979B:E02	ES 151
M00042981B:D11	ES 151
M00042983C:A11	ES 151
M00042983C:G06	ES 151
M00042986C:G12	ES 151
M00042988A:F06	ES 151
M00042997B:D06	ES 151
M00042998A:E03	ES 151
M00042998A:G04	ES 151
M00043001B:H10	ES 151
M00043001D:D03	ES 151
M00043002A:E05	ES 151
M00043003C:D08	ES 151
M00043011A:H12	ES 151
M00043015A:H10	ES 151
M00043022A:E12	ES 151
M00043026C:D07	ES 151
M00043028A:G05	ES 151
M00043029C:A06	ES 151
M00043032C:A10	ES 151
M00043034D:C01	ES 151
M00043036C:E05	ES 151
M00043036D:C09	ES 151
M00043040B:B07	ES 151
M00043044B:A12	ES 151
M00043044D:A09	ES 151
M00043045D:G12	ES 151
M00043046D:B11	ES 151

Clone Name	Tube
M00043060D:G12	ES 151
M00043066B:H11	ES 151
M00043067D:D10	ES 151
M00043125A:B11	ES 151
M00043125C:A11	ES 151
M00042611A:A06	ES 151
M00042611D:B12	ES 151
M00042612D:F06	ES 151
M00042614B:B05	ES 151
M00043073A:C12	ES 151
M00043078D:D04	ES 151
M00043081D:F05	ES 151
M00043087B:G07	ES 151
M00043093C:G11	ES 151
M00043095A:F09	ES 152
M00043096A:G04	ES 152
M00043108A:F06	ES 152
M00043109C:G01	ES 152
M00043131B:A09	ES 152
M00043133B:C11	ES 152
M00043138D:B11	ES 152
M00043143B:A10	ES 152
M00043148C:A09	ES 152
M00043154A:B07	ES 152
M00043162A:B08	ES 152
M00043162D:C12	ES 152
M00043164C:E12	ES 152
M00043165B:G01	ES 152
M00043173D:G03	ES 152
M00043184A:H08	ES 152
M00043187A:C04	ES 152
M00043191A:A07	ES 152
M00043192C:B12	ES 152
M00043200A:H09	ES 152
M00043200B:C08	ES 152
M00043202B:F01	ES 152
M00043203A:B09	ES 152
M00043210C:E05	ES 152
M00043211A:F01	ES 152
M00043213B:B12	ES 152
M00043215A:D02	ES 152
M00043220B:C04	ES 152
M00042591D:H03	ES 152
M00042592A:H10	ES 152
M00042593A:C02	ES 152

Clone Name	Tube
M00042593C:G06	ES 152
M00042595A:A11	ES 152
M00042595A:B01	ES 152
M00042596B:F06	ES 152
M00042596C:D07	ES 152
M00042597B:E12	ES 152
M00043416C:A02	ES 152
M00043417C:D05	ES 152
M00043418A:H10	ES 152
M00043419D:A10	ES 152
M00043428D:G08	ES 152
M00043430B:C02	ES 152
M00043431D:B08	ES 152
M00043433B:G09	ES 152
M00043433C:G07	ES 152
M00043437D:D04	ES 152
M00043440C:B07	ES 152
M00043446C:E12	ES 152
M00043447A:C07	ES 152
M00043449A:E12	ES 152
M00043450C:C06	ES 152
M00043453B:B09	ES 152
M00043458A:B12	ES 152
M00043461D:C02	ES 152
M00043461D:E06	ES 152
M00043465B:H02	ES 152
M00043465C:A03	ES 152
M00043465C:C09	ES 152
M00043476A:F07	ES 152
M00043483B:G10	ES 152
M00043491C:F04	ES 152
M00043492A:E01	ES 152
M00043513D:G08	ES 152
M00043516B:H09	ES 152
M00043518B:D06	ES 152
M00043526B:D10	ES 152
M00043527C:E09	ES 152
M00043528C:A02	ES 152
M00043616B:F02	ES 152
M00043616C:A05	ES 152
M00043632D:F09	ES 152
M00043634A:C10	ES 152
M00043635C:C11	ES 152
M00043636B:C06	ES 152
M00043637C:H01	ES 152

Clone Name	Tube
M00043638A:D06	ES 152
M00043640C:E03	ES 152
M00043648A:G07	ES 152
M00043649B:E07	ES 152
M00001338C:B02	ES 153
M00001338C:F05	ES 153
M00001338D:D01	ES 153
M00001340D:F07	ES 153
M00001344D:E08	ES 153
M00001346B:G11	ES 153
M00001348B:B03	ES 153
M00001349C:B04	ES 153
M00001351B:E11	ES 153
M00001352B:B02	ES 153
M00001353A:H07	ES 153
M00001353C:A05	ES 153
M00001353D:E05	ES 153
M00001356D:E06	ES 153
M00001358A:E08	ES 153
M00001359A:H10	ES 153
M00001361A:C12	ES 153
M00001361B:A12	ES 153
M00001362A:F09	ES 153
M00001364A:C09	ES 153
M00001364C:H10	ES 153
M00001368A:A08	ES 153
M00001368A:B07	ES 153
M00001368A:C02	ES 153
M00001369A:G06	ES 153
M00001374A:B02	ES 153
M00001374C:B10	ES 153
M00001375B:D04	ES 153
M00001378C:E10	ES 153
M00001379A:F09	ES 153
M00001382D:A07	ES 153
M00001382D:H08	ES 153
M00001384A:A07	ES 153
M00001385A:E07	ES 153
M00001386B:F11	ES 153
M00001387A:C12	ES 153
M00001387B:A11	ES 153
M00001389B:E10	ES 153
M00001389D:D06	ES 153
M00001390D:E02	ES 153
M00001391D:D03	ES 153

Clone Name	Tube
M00001393B:C03	ES 153
M00001393C:E08	ES 153
M00001393C:F04	ES 153
M00001393D:E02	ES 153
M00001396B:B01	ES 153
M00001396B:B12	ES 153
M00001396D:H02	ES 153
M00001397C:H08	ES 153
M00001399B:B01	ES 153
M00001399C:A01	ES 153
M00001403C:B03	ES 153
M00001403D:C12	ES 153
M00001406B:H09	ES 153
M00001406D:F06	ES 153
M00001410A:G10	ES 153
M00001416B:A05	ES 153
M00001421B:E07	ES 153
M00001422B:D06	ES 153
M00001424B:H06	ES 153
M00001424D:D02	ES 153
M00001426C:F06	ES 153
M00001428B:C10	ES 153
M00001429B:G05	ES 153
M00001430B:C01	ES 153
M00001433B:E02	ES 153
M00001442A:F08	ES 153
M00001442C:G12	ES 153
M00001444B:E04	ES 153
M00001444C:D11	ES 153
M00001445B:F06	ES 153
M00001449B:H10	ES 153
M00001451C:E10	ES 153
M00001460C:E10	ES 153
M00001461D:B10	ES 153
M00001461D:C10	ES 153
M00001465C:A02	ES 153
M00001466B:F03	ES 153
M00001467C:D04	ES 153
M00001477D:G09	ES 153
M00001485C:F06	ES 153
M00001488C:A03	ES 153
M00001497C:F10	ES 153
M00001503B:H10	ES 153
M00001506B:D11	ES 153
M00001512D:F08	ES 153

Clone Name	Tube
M00001518B:D10	ES 153
M00001528C:C03	ES 153
M00001532A:G08	ES 153
M00001533C:G11	ES 153
M00001533D:A01	ES 153
M00001534C:E07	ES 153
M00001535B:B10	ES 153
M00001535B:E02	ES 153
M00001537B:H10	ES 153
M00001538B:A07	ES 153
M00001539C:F12	ES 154
M00001542B:F09	ES 154
M00001543C:A08	ES 154
M00001544B:B05	ES 154
M00001544B:E06	ES 154
M00001546B:C11	ES 154
M00001548B:D06	ES 154
M00001550A:H06	ES 154
M00001550D:B11	ES 154
M00001551D:D01	ES 154
M00001551D:H09	ES 154
M00001554C:G10	ES 154
M00001558A:E06	ES 154
M00001559A:H09	ES 154
M00001561D:H04	ES 154
M00001562B:B02	ES 154
M00001562D:B07	ES 154
M00001565A:H05	ES 154
M00001568C:A03	ES 154
M00001570A:B07	ES 154
M00001591B:H05	ES 154
M00001596A:D02	ES 154
M00001600B:G01	ES 154
M00001605B:B05	ES 154
M00001606B:A10	ES 154
M00001606D:D06	ES 154
M00001607A:E04	ES 154
M00001607D:H09	ES 154
M00001609D:C11	ES 154
M00001616D:F03	ES 154
M00001617C:F10	ES 154
M00001618C:D01	ES 154
M00001619C:H09	ES 154
M00001620B:A03	ES 154
M00001623D:A10	ES 154

Clone Name	Tube
M00001623D:E12	ES 154
M00001624A:C01	ES 154
M00001625D:B04	ES 154
M00001626A:D07	ES 154
M00001632C:A10	ES 154
M00001633D:C11	ES 154
M00001637D:C12	ES 154
M00001648A:D10	ES 154
M00001661D:F06	ES 154
M00001663A:A12	ES 154
M00001671A:H10	ES 154
M00001671C:F03	ES 154
M00001675B:D06	ES 154
M00001677B:H08	ES 154
M00001680A:A01	ES 154
M00001683B:F11	ES 154
M00001684D:E04	ES 154
M00001686B:H01	ES 154
M00001686D:F06	ES 154
M00001688B:B11	ES 154
M00001692C:C04	ES 154
M00001771B:E06	ES 154
M00003746C:E11	ES 154
M00003749C:C08	ES 154
M00003753A:C11	ES 154
M00003758B:D07	ES 154
M00003758B:F06	ES 154
M00003760C:G10	ES 154
M00003761B:B02	ES 154
M00003763A:B02	ES 154
M00003763B:B10	ES 154
M00003764A:H09	ES 154
M00003764B:F11	ES 154
M00003764B:H11	ES 154
M00003764D:F07	ES 154
M00003768D:D08	ES 154
M00003770C:A10	ES 154
M00003771D:A03	ES 154
M00003773A:F10	ES 154
M00003780A:G01	ES 154
M00003782A:B02	ES 154
M00003785D:F07	ES 154
M00003787D:A10	ES 154
M00003808A:F11	ES 154
M00003808B:E07	ES 154

Clone Name	Tube
M00003812C:A03	ES 154
M00003814A:G05	ES 154
M00003819B:B01	ES 154
M00003820B:F11	ES 154
M00003821C:E12	ES 154
M00003822C:A09	ES 154
M00003822D:A02	ES 154
M00003823B:A06	ES 154
M00003825A:H10	ES 154
M00003828A:D11	ES 154
M00003830B:C06	ES 154
M00003830C:D02	ES 154
M00003837C:D10	ES 154
M00003839C:H10	ES 154
M00003842D:D11	ES 154
M00003842D:H09	ES 154
M00003845A:C07	ES 155
M00003845D:G03	ES 155
M00003847A:H04	ES 155
M00003848C:G09	ES 155
M00003851B:A01	ES 155
M00003854B:F07	ES 155
M00003855C:F02	ES 155
M00003884A:E12	ES 155
M00003887C:E09	ES 155
M00003888B:F09	ES 155
M00003891B:H02	ES 155
M00003898C:A01	ES 155
M00003900C:D12	ES 155
M00003906A:C02	ES 155
M00003911C:A09	ES 155
M00003914A:A08	ES 155
M00003915C:D10	ES 155
M00003915C:G08	ES 155
M00003916A:E04	ES 155
M00003926A:F11	ES 155
M00003935B:B01	ES 155
M00003938C:A05	ES 155
M00003942A:D01	ES 155
M00003958C:H08	ES 155
M00003959D:A05	ES 155
M00003960D:C12	ES 155
M00003963D:F01	ES 155
M00003965D:D11	ES 155
M00003968C:G03	ES 155

Clone Name	Tube
M00003970D:H07	ES 155
M00003972C:F07	ES 155
M00003974C:E11	ES 155
M00003974D:E02	ES 155
M00003979B:A04	ES 155
M00003980D:C06	ES 155
M00003985D:B02	ES 155
M00003988D:B01	ES 155
M00003991A:C11	ES 155
M00003993C:D07	ES 155
M00003993D:B03	ES 155
M00003994A:B10	ES 155
M00003996B:H07	ES 155
M00003998B:G10	ES 155
M00004028B:F10	ES 155
M00004029D:A01	ES 155
M00004031C:G06	ES 155
M00004036B:A11	ES 155
M00004036D:C12	ES 155
M00004038A:A04	ES 155
M00004042B:A11	ES 155
M00004047C:B09	ES 155
M00004047D:F12	ES 155
M00004053D:F09	ES 155
M00004054A:D03	ES 155
M00004055C:B10	ES 155
M00004055D:D05	ES 155
M00004057D:G01	ES 155
M00004061B:E05	ES 155
M00004062D:A02	ES 155
M00004066D:G10	ES 155
M00004067B:D03	ES 155
M00004080C:C04	ES 155
M00004085A:H01	ES 155
M00004085B:H02	ES 155
M00004087C:E02	ES 155
M00004093A:C03	ES 155
M00004096D:F02	ES 155
M00004102A:E03	ES 155
M00004103C:E10	ES 155
M00004104A:A12	ES 155
M00004110D:F09	ES 155
M00004114C:D11	ES 155
M00004115A:G12	ES 155
M00004118C:D12	ES 155

Clone Name	Tube
M00004122C:D01	ES 155
M00004134A:A08	ES 155
M00004136C:B12	ES 155
M00004139B:F01	ES 155
M00004141A:D01	ES 155
M00004141B:B01	ES 155
M00004141B:F08	ES 155
M00004143B:B04	ES 155
M00004144D:B02	ES 155
M00004146A:C11	ES 155
M00004146B:E08	ES 155
M00004146C:B04	ES 155
M00004147C:E01	ES 155
M00004151B:A07	ES 155
M00004155A:H03	ES 155
M00004155C:A10	ES 155
M00004158B:E03	ES 155
M00004158D:E08	ES 155
M00004159C:D10	ES 155
M00004159D:F12	ES 155
M00004160D:F06	ES 155
M00004160D:G05	ES 155
M00004162D:F02	ES 156
M00004163B:C03	ES 156
M00004163C:A03	ES 156
M00004164B:E12	ES 156
M00004165C:A11	ES 156
M00004166C:B10	ES 156
M00004169A:E04	ES 156
M00004170A:F03	ES 156
M00004171B:B03	ES 156
M00004172C:A08	ES 156
M00004172D:B12	ES 156
M00004172D:F04	ES 156
M00004175D:E06	ES 156
M00004176C:A09	ES 156
M00004179C:B06	ES 156
M00004179D:A12	ES 156
M00004187B:C02	ES 156
M00004189A:C12	ES 156
M00004192C:B06	ES 156
M00004195A:F07	ES 156
M00004200C:A04	ES 156
M00004201D:C01	ES 156
M00004201D:C03	ES 156

Clone Name	Tube
M00004204C:H08	ES 156
M00004207C:A04	ES 156
M00004208A:D08	ES 156
M00004210A:A03	ES 156
M00004212D:C03	ES 156
M00004214A:E05	ES 156
M00004214D:A05	ES 156
M00004215B:C05	ES 156
M00004220D:C11	ES 156
M00004225D:E03	ES 156
M00004229B:B06	ES 156
M00004230D:B05	ES 156
M00004237C:D10	ES 156
M00004242D:H01	ES 156
M00004245C:G10	ES 156
M00004246B:H07	ES 156
M00004251D:D03	ES 156
M00004263C:D03	ES 156
M00004266B:F07	ES 156
M00004269A:F11	ES 156
M00004269A:G11	ES 156
M00004269B:B04	ES 156
M00004270A:E09	ES 156
M00004276C:A08	ES 156
M00004277D:B02	ES 156
M00004278A:G06	ES 156
M00004278C:B10	ES 156
M00004281A:C04	ES 156
M00004282A:D01	ES 156
M00004282B:D07	ES 156
M00004282C:A12	ES 156
M00004284A:F08	ES 156
M00004295D:C07	ES 156
M00004296B:D03	ES 156
M00004303C:C05	ES 156
M00004310B:E02	ES 156
M00004316A:B03	ES 156
M00004320C:E07	ES 156
M00004321C:C11	ES 156
M00004322B:D03	ES 156
M00004324A:B03	ES 156
M00004324A:D10	ES 156
M00004324A:D05	ES 156
M00004328A:D01	ES 156
M00004330A:A01	ES 156

Clone Name	Tube
M00004336A:A01	ES 156
M00004341C:A09	ES 156
M00004341C:E05	ES 156
M00004344A:G11	ES 156
M00004344D:C12	ES 156
M00004347B:E04	ES 156
M00004347C:A05	ES 156
M00004350A:A04	ES 156
M00004351B:G07	ES 156
M00004352A:D08	ES 156
M00004357B:B06	ES 156
M00004358B:G02	ES 156
M00004359A:E01	ES 156
M00004360C:D09	ES 156
M00004365C:C09	ES 156
M00004365C:G11	ES 156
M00004366D:C11	ES 156
M00004368A:B11	ES 156
M00004372A:E12	ES 156
M00004376D:A12	ES 156
M00004385C:H12	ES 156
M00004393C:D06	ES 156
M00004406A:G09	ES 156
M00004416B:G10	ES 156
M00004418B:A11	ES 156
M00004419A:G02	ES 156
M00004420D:E05	ES 156
M00004430A:A05	ES 156
M00004430B:B10	ES 157
M00004443C:F07	ES 157
M00004462D:D12	ES 157
M00004502A:D12	ES 157
M00004507D:E03	ES 157
M00004509B:B10	ES 157
M00004509D:C06	ES 157
M00004603B:E02	ES 157
M00004603C:C10	ES 157
M00004606D:H09	ES 157
M00004608A:C10	ES 157
M00004608A:H04	ES 157
M00004609A:E09	ES 157
M00023389A:G04	ES 157
M00023394D:D10	ES 157
M00026809A:H08	ES 157
M00026818C:E01	ES 157

Clone Name	Tube
M00026836B:H03	ES 157
M00026842B:A01	ES 157
M00026842D:C02	ES 157
M00026850B:C09	ES 157
M00026856B:G03	ES 157
M00026900A:H07	ES 157
M00026907D:E07	ES 157
M00026910B:G06	ES 157
M00026914C:H09	ES 157
M00026936D:C07	ES 157
M00026961A:B06	ES 157
M00026994D:D07	ES 157
M00027004C:C11	ES 157
M00027017A:B09	ES 157
M00027036A:B06	ES 157
M00027050A:B02	ES 157
M00027052A:E10	ES 157
M00027057C:D10	ES 157
M00027064B:D06	ES 157
M00027081A:A08	ES 157
M00027093A:H02	ES 157
M00027131A:B03	ES 157
M00027159C:F07	ES 157
M00027167C:B10	ES 157
M00027168B:H08	ES 157
M00027170D:C07	ES 157
M00027173C:E11	ES 157
M00027177B:D04	ES 157
M00027178B:A11	ES 157
M00027182B:G06	ES 157
M00027189C:B10	ES 157
M00027193C:A07	ES 157
M00027215A:F06	ES 157
M00027215B:B12	ES 157
M00027244C:B06	ES 157
M00027247C:D02	ES 157
M00027262A:A07	ES 157
M00027270A:D04	ES 157
M00027274A:A09	ES 157
M00027290C:F06	ES 157
M00027291A:G08	ES 157
M00027311A:H09	ES 157
M00027313C:E01	ES 157
M00027314D:E02	ES 157
M00027316C:C03	ES 157

Clone Name	Tube
M00027319C:C03	ES 157
M00027319D:F07	ES 157
M00027332B:H09	ES 157
M00027359B:A06	ES 157
M00027363D:G04	ES 157
M00027364B:E12	ES 157
M00027376C:A02	ES 157
M00027381B:B04	ES 157
M00027400D:H02	ES 157
M00027433B:D12	ES 157
M00027457B:E11	ES 157
M00027459C:B10	ES 157
M00027467A:C07	ES 157
M00027475D:A01	ES 157
M00027480C:E09	ES 157
M00027485C:F07	ES 157
M00027506B:G01	ES 157
M00027513D:F06	ES 157
M00027523A:H05	ES 157
M00027527B:C05	ES 157
M00027549C:G03	ES 157
M00027569A:E05	ES 157
M00027586A:C09	ES 157
M00027589B:G07	ES 157
M00027591A:E04	ES 157
M00027600B:C07	ES 157
M00027605B:D09	ES 157
M00027688C:C01	ES 157
M00027717C:C06	ES 157
M00027724D:D04	ES 157
M00027734D:C03	ES 157
M00027746A:D06	ES 157
M00027801B:D07	ES 157
M00027806C:H05	ES 157
M00028055B:G07	ES 158
M00028063C:H01	ES 158
M00028067A:C11	ES 158
M00028069D:H02	ES 158
M00028070A:H09	ES 158
M00028070D:C03	ES 158
M00028188C:H11	ES 158
M00028193B:E07	ES 158
M00028196A:G03	ES 158
M00028210B:H03	ES 158
M00028211A:F10	ES 158

Clone Name	Tube
M00028212D:C05	ES 158
M00028219B:H05	ES 158
M00028361B:H08	ES 158
M00028366B:B08	ES 158
M00028616C:D09	ES 158
M00028620C:C07	ES 158
M00028763A:G11	ES 158
M00028764B:D03	ES 158
M00028771A:E02	ES 158
M00028773C:C05	ES 158
M00028774D:E10	ES 158
M00028777B:G04	ES 158
M00028782A:F01	ES 158
M00028784A:D12	ES 158
M00028786B:A04	ES 158
M00031370B:C01	ES 158
M00031416D:H05	ES 158
M00031484A:D03	ES 158
M00031485B:G05	ES 158
M00032471D:A05	ES 158
M00032473B:A03	ES 158
M00032474A:G03	ES 158
M00032475A:A06	ES 158
M00032489B:G12	ES 158
M00032490D:E08	ES 158
M00032494C:H08	ES 158
M00032497D:B10	ES 158
M00032504B:B10	ES 158
M00032507D:G08	ES 158
M00032508A:E03	ES 158
M00032515A:B12	ES 158
M00032517C:E10	ES 158
M00032519D:F08	ES 158
M00032534B:E12	ES 158
M00032541C:G03	ES 158
M00032553A:A07	ES 158
M00032556D:A03	ES 158
M00032562C:F01	ES 158
M00032567B:G05	ES 158
M00032568B:F08	ES 158
M00032577D:F01	ES 158
M00032580D:A09	ES 158
M00032581B:A09	ES 158
M00032584A:D06	ES 158
M00032586C:B04	ES 158

Clone Name	Tube
M00032590B:H01	ES 158
M00032594C:F05	ES 158
M00032597A:H02	ES 158
M00032605B:D09	ES 158
M00032613A:E11	ES 158
M00032614C:B10	ES 158
M00032614D:D08	ES 158
M00032620B:F06	ES 158
M00032621A:F11	ES 158
M00032628C:B06	ES 158
M00032634B:D09	ES 158
M00032637A:F09	ES 158
M00032638B:F02	ES 158
M00032644C:B05	ES 158
M00032645D:C01	ES 158
M00032647B:F06	ES 158
M00032652C:C07	ES 158
M00032666A:C02	ES 158
M00032671B:D06	ES 158
M00032671B:D08	ES 158
M00032676C:C10	ES 158
M00032688C:A03	ES 158
M00032700A:E09	ES 158
M00032707D:F08	ES 158
M00032711B:F01	ES 158
M00032723D:H02	ES 158
M00032727A:E04	ES 158
M00032728D:F01	ES 158
M00032729A:F10	ES 158
M00032733B:F12	ES 158
M00032734B:E12	ES 158
M00032734C:C05	ES 158
M00032749D:G03	ES 158
M00032753A:C07	ES 158
M00032759A:A03	ES 158
M00032765A:C05	ES 158
M00032770C:G11	ES 158
M00032772D:D03	ES 158
M00032773D:F08	ES 158
M00032774C:C04	ES 158
M00032787D:C05	ES 159
M00032791B:H11	ES 159
M00032791D:F01	ES 159
M00032792C:B01	ES 159
M00032793A:G06	ES 159

Clone Name	Tube
M00032795C:A03	ES 159
M00032797D:D08	ES 159
M00032825B:F08	ES 159
M00032826C:D10	ES 159
M00032828A:A06	ES 159
M00032829D:A05	ES 159
M00032830D:D02	ES 159
M00032831A:C07	ES 159
M00032831A:E09	ES 159
M00032835D:G04	ES 159
M00032836B:A07	ES 159
M00032848D:B10	ES 159
M00032892C:C12	ES 159
M00032908A:D08	ES 159
M00032918D:B04	ES 159
M00032928C:D02	ES 159
M00032944A:B07	ES 159
M00032945D:B07	ES 159
M00032979D:C11	ES 159
M00032979D:H07	ES 159
M00032985D:G09	ES 159
M00032987B:F01	ES 159
M00032988B:G01	ES 159
M00033006A:F10	ES 159
M00033028C:A02	ES 159
M00033028D:C10	ES 159
M00033037B:F04	ES 159
M00033041A:B11	ES 159
M00033055D:D02	ES 159
M00033071C:G05	ES 159
M00033071D:E08	ES 159
M00033072A:A09	ES 159
M00033080C:A07	ES 159
M00033081D:D11	ES 159
M00033144A:D02	ES 159
M00033146D:A03	ES 159
M00033147C:B08	ES 159
M00033149B:E10	ES 159
M00033150B:E02	ES 159
M00033150C:A11	ES 159
M00033183B:F10	ES 159
M00033218C:F07	ES 159
M00033223C:G04	ES 159
M00033230C:G10	ES 159
M00033232B:C08	ES 159

Clone Name	Tube
M00033246A:H12	ES 159
M00033248D:H11	ES 159
M00033264B:E06	ES 159
M00033274D:F03	ES 159
M00033311B:G10	ES 159
M00033324B:F04	ES 159
M00033326B:B05	ES 159
M00033329C:C02	ES 159
M00033296C:C11	ES 159
M00033302A:E11	ES 159
M00033302B:F10	ES 159
M00033303C:F09	ES 159
M00033342B:F03	ES 159
M00033344A:B06	ES 159
M00033359C:H05	ES 159
M00033360C:A03	ES 159
M00033374D:C07	ES 159
M00033413A:A08	ES 159
M00033420B:E08	ES 159
M00033434D:F05	ES 159
M00033441A:B12	ES 159
M00033445D:G03	ES 159
M00038290A:D12	ES 159
M00038304B:E02	ES 159
M00038389D:D10	ES 159
M00038390B:F02	ES 159
M00038616C:C09	ES 159
M00038616D:B07	ES 159
M00038618D:D08	ES 159
M00038619B:F09	ES 159
M00038619D:C12	ES 159
M00039001A:B10	ES 159
M00039024D:E12	ES 159
M00039055C:A01	ES 159
M00039056B:G01	ES 159
M00039063C:H09	ES 159
M00039067A:C05	ES 159
M00039067B:F07	ES 159
M00039076D:G04	ES 159
M00039078B:B03	ES 159
M00039078D:C10	ES 159
M00039081B:C04	ES 159
M00039081B:G07	ES 159
M00039100A:G04	ES 159
M00039105D:A08	ES 159

Clone Name	Tube
M00039107A:E12	ES 159
M00039111A:C12	ES 160
M00039121D:E07	ES 160
M00039124D:H01	ES 160
M00039125D:H12	ES 160
M00039131C:B09	ES 160
M00039133B:D06	ES 160
M00039133C:F12	ES 160
M00039134D:F08	ES 160
M00039138B:G05	ES 160
M00039140A:F05	ES 160
M00039143A:F04	ES 160
M00039143D:C10	ES 160
M00039146B:G04	ES 160
M00039162D:C04	ES 160
M00039165D:C04	ES 160
M00039175A:F01	ES 160
M00039204A:E09	ES 160
M00039207A:F07	ES 160
M00039219B:C08	ES 160
M00039222B:A04	ES 160
M00039225A:D11	ES 160
M00039246B:A08	ES 160
M00039248C:A08	ES 160
M00039251C:H12	ES 160
M00039251D:B08	ES 160
M00039255D:B01	ES 160
M00039258C:C01	ES 160
M00039270D:D02	ES 160
M00039275B:E02	ES 160
M00039278C:D03	ES 160
M00039284D:H07	ES 160
M00039285B:G04	ES 160
M00039291D:F02	ES 160
M00039294C:B09	ES 160
M00039302B:E10	ES 160
M00039326A:G07	ES 160
M00039326C:B08	ES 160
M00039331B:F09	ES 160
M00039338B:F07	ES 160
M00039344C:A11	ES 160
M00039349D:B11	ES 160
M00039381C:C07	ES 160
M00039383A:H07	ES 160
M00039411D:D09	ES 160

Clone Name	Tube
M00039413C:E06	ES 160
M00039430A:E04	ES 160
M00039433B:D06	ES 160
M00039433C:E03	ES 160
M00039438B:D08	ES 160
M00039440C:G06	ES 160
M00039457D:C02	ES 160
M00039471D:G10	ES 160
M00039472B:E05	ES 160
M00039478C:B02	ES 160
M00039554D:B09	ES 160
M00039556C:G05	ES 160
M00039559B:C07	ES 160
M00039560B:G09	ES 160
M00039560C:G06	ES 160
M00039617C:A10	ES 160
M00039654C:C11	ES 160
M00039668C:F01	ES 160
M00039672C:D05	ES 160
M00039686C:C01	ES 160
M00039694C:H01	ES 160
M00039698C:B03	ES 160
M00039710B:A01	ES 160
M00039710B:E01	ES 160
M00039785C:H12	ES 160
M00039786D:A10	ES 160
M00039805B:B06	ES 160
M00039806B:D05	ES 160
M00039820B:F06	ES 160
M00039822A:H02	ES 160
M00039826B:F09	ES 160
M00039826D:E04	ES 160
M00039828B:H06	ES 160
M00039829B:E01	ES 160
M00039860B:E01	ES 160
M00039860D:B02	ES 160
M00039861C:B12	ES 160
M00039865A:C09	ES 160
M00039869A:H01	ES 160
M00039871C:G05	ES 160
M00039873B:H04	ES 160
M00039874A:B06	ES 160
M00039885C:D11	ES 160
M00039894C:D09	ES 160
M00039895D:C04	ES 160

Clone Name	Tube
M00039900B:G04	ES 160
M00039915B:E08	ES 160
M00039921A:B10	ES 160
M00004824A:D12	ES 160
M00004824D:H05	ES 160
M00004831C:G11	ES 160
M00004832D:G04	ES 160
M00004836B:C02	ES 161
M00004839B:C12	ES 161
M00004843A:G12	ES 161
M00004846A:A10	ES 161
M00004850A:B02	ES 161
M00004852D:C06	ES 161
M00004856D:F09	ES 161
M00004873B:G04	ES 161
M00004876B:A06	ES 161
M00005002A:C03	ES 161
M00005003D:C02	ES 161
M00005013D:H05	ES 161
M00005014B:F02	ES 161
M00005016C:E04	ES 161
M00005309B:A11	ES 161
M00005314A:G10	ES 161
M00005332A:C06	ES 161
M00005333D:D08	ES 161
M00005346D:A03	ES 161
M00005349C:C02	ES 161
M00005359B:B08	ES 161
M00005359B:D09	ES 161
M00005364B:E10	ES 161
M00005365A:F05	ES 161
M00005366D:F08	ES 161
M00005367D:A11	ES 161
M00005375D:A10	ES 161
M00005379A:D10	ES 161
M00005380B:H10	ES 161
M00005383A:C11	ES 161
M00005385A:B12	ES 161
M00005385D:F07	ES 161
M00005387A:B03	ES 161
M00005392A:G06	ES 161
M00005401D:F09	ES 161
M00005403C:A01	ES 161
M00005405C:D01	ES 161
M00005409D:B02	ES 161

Clone Name	Tube
M00005413D:A05	ES 161
M00005422B:B08	ES 161
M00005422D:H02	ES 161
M00005422D:H10	ES 161
M00005423A:C11	ES 161
M00005423C:A10	ES 161
M00005423C:D07	ES 161
M00005434A:C03	ES 161
M00005442A:B10	ES 161
M00005445A:E07	ES 161
M00005445D:D04	ES 161
M00005445D:F11	ES 161
M00005452B:G03	ES 161
M00005452D:E05	ES 161
M00005460D:C11	ES 161
M00005461A:D12	ES 161
M00005463A:G02	ES 161
M00005466C:B01	ES 161
M00005468A:C04	ES 161
M00005468D:C01	ES 161
M00005474C:H09	ES 161
M00005485C:H04	ES 161
M00005489B:C08	ES 161
M00005500A:D04	ES 161
M00005504C:F12	ES 161
M00005504D:F06	ES 161
M00005505A:F01	ES 161
M00005505B:E01	ES 161
M00005506C:E09	ES 161
M00005506D:E11	ES 161
M00005507B:A03	ES 161
M00005511A:F05	ES 161
M00005512B:H01	ES 161
M00005515D:F02	ES 161
M00005520B:E01	ES 161
M00005520B:H05	ES 161
M00005524C:H04	ES 161
M00005535B:B01	ES 161
M00005540A:F09	ES 161
M00005557D:H10	ES 161
M00005569D:G09	ES 161
M00005570A:B08	ES 161
M00005570A:D05	ES 161
M00005603B:H03	ES 161
M00005606D:B12	ES 161

Clone Name	Tube
M00005607B:C04	ES 161
M00005616B:F07	ES 161
M00005622A:H02	ES 161
M00005623B:G01	ES 161
M00005626D:G11	ES 161
M00005634A:F07	ES 161
M00005641B:E09	ES 161
M00005643D:A05	ES 161
M00005674C:F04	ES 161
M00005675D:D09	ES 161
M00005689C:B02	ES 161
M00005703B:E03	ES 161
M00005703D:G10	ES 161
M00005710B:H03	ES 162
M00005743D:A12	ES 162
M00005763D:A01	ES 162
M00005766D:D12	ES 162
M00005771D:C02	ES 162
M00005819D:F09	ES 162
M00005822C:A04	ES 162
M00006576D:C02	ES 162
M00006577A:H10	ES 162
M00006582D:A09	ES 162
M00006585A:D07	ES 162
M00006585A:F09	ES 162
M00006586D:D04	ES 162
M00006592A:A12	ES 162
M00006595B:C10	ES 162
M00006601D:G05	ES 162
M00006631C:A04	ES 162
M00006631D:D02	ES 162
M00006636B:E04	ES 162
M00006641B:F05	ES 162
M00006646A:A07	ES 162
M00006678A:A03	ES 162
M00006678C:C02	ES 162
M00006712C:H01	ES 162
M00006714C:D06	ES 162
M00006738A:F12	ES 162
M00006739B:A04	ES 162
M00006740B:A09	ES 162
M00006743A:D04	ES 162
M00006743A:H11	ES 162
M00006756B:G06	ES 162
M00006756C:A02	ES 162

Clone Name	Tube
M00006861D:H10	ES 162
M00006872D:B07	ES 162
M00006877B:C09	ES 162
M00006877C:F11	ES 162
M00006884D:A08	ES 162
M00006885A:F07	ES 162
M00006890C:F10	ES 162
M00006904D:A02	ES 162
M00006907A:C09	ES 162
M00006907B:C06	ES 162
M00006989B:G05	ES 162
M00006994C:F06	ES 162
M00007002C:A10	ES 162
M00007006C:C12	ES 162
M00007007A:E04	ES 162
M00007031A:E02	ES 162
M00007032A:B05	ES 162
M00007032C:A12	ES 162
M00007046D:C09	ES 162
M00007048B:E11	ES 162
M00007048C:A12	ES 162
M00007059B:D07	ES 162
M00007060D:G07	ES 162
M00007064D:D12	ES 162
M00007070C:C01	ES 162
M00007081B:C08	ES 162
M00007081B:E09	ES 162
M00007082D:E05	ES 162
M00007098A:E10	ES 162
M00007103C:C12	ES 162
M00007103D:C02	ES 162
M00007112D:D03	ES 162
M00007117A:C11	ES 162
M00007126A:A02	ES 162
M00007141C:B05	ES 162
M00007154A:E06	ES 162
M00007155C:D07	ES 162
M00007155D:C09	ES 162
M00007158D:D03	ES 162
M00007178A:C02	ES 162
M00007195C:E11	ES 162
M00007197B:B05	ES 162
M00007202B:F01	ES 162
M00007947A:B06	ES 162
M00007953D:F07	ES 162

Clone Name	Tube
M00007969D:C01	ES 162
M00007973B:D11	ES 162
M00007975C:A10	ES 162
M00007975D:F12	ES 162
M00007980A:B01	ES 162
M00007980B:A07	ES 162
M00007981C:F07	ES 162
M00007985C:D08	ES 162
M00008001B:F05	ES 162
M00008007B:E03	ES 162
M00008016B:E09	ES 162
M00008019B:A01	ES 162
M00008020D:D05	ES 162
M00008020D:F02	ES 162
M00008021C:G12	ES 162
M00008045C:A05	ES 162
M00008055D:G03	ES 162
M00008059B:F08	ES 162
M00008059D:B08	ES 162
M00008065D:A07	ES 163
M00008071D:H03	ES 163
M00008073A:D01	ES 163
M00008073D:D01	ES 163
M00021649B:A02	ES 163
M00021650D:A11	ES 163
M00021653A:B02	ES 163
M00021668D:A03	ES 163
M00021676C:G03	ES 163
M00021677A:D09	ES 163
M00021678A:H03	ES 163
M00021678D:H04	ES 163
M00021681C:C09	ES 163
M00021690A:C03	ES 163
M00021697C:B07	ES 163
M00021700D:H03	ES 163
M00021852C:H02	ES 163
M00021855D:F10	ES 163
M00021866C:H08	ES 163
M00021896D:A05	ES 163
M00021923A:B12	ES 163
M00021923D:H02	ES 163
M00021933B:F02	ES 163
M00021941A:D09	ES 163
M00021952B:G06	ES 163
M00021958B:E08	ES 163

Clone Name	Tube
M00021967D:H06	ES 163
M00021971C:B11	ES 163
M00021974D:F01	ES 163
M00021981A:C02	ES 163
M00021991D:F09	ES 163
M00021998B:D09	ES 163
M00022009C:A08	ES 163
M00022016B:F01	ES 163
M00022032A:G05	ES 163
M00022051B:D07	ES 163
M00022069D:C12	ES 163
M00022070B:B04	ES 163
M00022073C:C07	ES 163
M00022081A:B07	ES 163
M00022088B:F10	ES 163
M00022088B:H02	ES 163
M00022088D:E10	ES 163
M00022090B:A10	ES 163
M00022092D:A11	ES 163
M00022094B:G02	ES 163
M00022096D:A03	ES 163
M00022103C:D05	ES 163
M00022104A:G08	ES 163
M00022117C:A02	ES 163
M00022118A:E06	ES 163
M00022140D:A07	ES 163
M00022144C:E12	ES 163
M00022158B:B09	ES 163
M00022170C:C01	ES 163
M00022171A:F03	ES 163
M00022185A:B03	ES 163
M00022193B:A09	ES 163
M00022193C:C09	ES 163
M00022200B:B05	ES 163
M00022202C:C04	ES 163
M00022208B:D03	ES 163
M00022208C:E04	ES 163
M00022208C:F08	ES 163
M00022212D:G02	ES 163
M00022216D:D10	ES 163
M00022218B:B12	ES 163
M00022220A:A07	ES 163
M00022224A:C07	ES 163
M00022224A:G07	ES 163
M00022228B:B11	ES 163

Clone Name	Tube
M00022229D:E01	ES 163
M00022237C:E04	ES 163
M00022237D:D06	ES 163
M00022238C:G04	ES 163
M00022240B:C12	ES 163
M00022240D:B11	ES 163
M00022249D:C01	ES 163
M00022250A:B04	ES 163
M00022262A:F06	ES 163
M00022262B:B06	ES 163
M00022264A:B02	ES 163
M00022265A:F11	ES 163
M00022269C:A04	ES 163
M00022273A:E03	ES 163
M00022282B:C09	ES 163
M00022305A:B04	ES 163
M00022363C:D05	ES 163
M00022367D:G11	ES 163
M00022368A:B11	ES 163
M00022372D:H12	ES 163
M00022374C:E11	ES 163
M00022376D:D05	ES 163
M00022383C:A12	ES 163
M00022386D:F10	ES 163
M00022392B:F01	ES 163
M00022403C:E12	ES 164
M00022415C:D12	ES 164
M00022416D:D01	ES 164
M00022421A:F12	ES 164
M00022425A:C09	ES 164
M00022430C:C06	ES 164
M00022435B:G12	ES 164
M00022436C:F11	ES 164
M00022438C:H09	ES 164
M00022442B:G03	ES 164
M00022446C:H06	ES 164
M00022449D:F08	ES 164
M00022452B:E06	ES 164
M00022454C:B08	ES 164
M00022457A:G05	ES 164
M00022467D:B03	ES 164
M00022470D:B02	ES 164
M00022472D:B01	ES 164
M00022474B:C08	ES 164
M00022475D:C07	ES 164

Clone Name	Tube
M00022481B:A04	ES 164
M00022485B:E07	ES 164
M00022487B:A08	ES 164
M00022487C:C02	ES 164
M00022491A:A08	ES 164
M00022491D:A10	ES 164
M00022494B:D06	ES 164
M00022494D:A05	ES 164
M00022499D:D08	ES 164
M00022507C:C08	ES 164
M00022509A:H02	ES 164
M00022509B:D11	ES 164
M00022512B:A09	ES 164
M00022516B:C05	ES 164
M00022525B:D09	ES 164
M00022530B:C04	ES 164
M00022537B:C06	ES 164
M00022546B:E05	ES 164
M00022559D:G10	ES 164
M00022563B:C08	ES 164
M00022590B:E05	ES 164
M00022600D:B05	ES 164
M00022601B:G06	ES 164
M00022618B:D09	ES 164
M00022618C:E04	ES 164
M00022627B:H03	ES 164
M00022634A:C07	ES 164
M00022634B:H09	ES 164
M00022638A:D03	ES 164
M00022642A:G08	ES 164
M00022648A:D08	ES 164
M00022656D:D07	ES 164
M00022662C:H04	ES 164
M00022662D:H03	ES 164
M00022672C:H04	ES 164
M00022674C:H08	ES 164
M00022677C:C01	ES 164
M00022678B:C08	ES 164
M00022681D:E10	ES 164
M00022682D:A10	ES 164
M00022684A:E06	ES 164
M00022690A:A07	ES 164
M00022694A:F05	ES 164
M00022696B:C11	ES 164
M00039921C:H11	ES 164

Clone Name	Tube
M00039929B:E06	ES 164
M00039929D:H10	ES 164
M00039932B:A07	ES 164
M00039976C:F11	ES 164
M00039977B:D12	ES 164
M00039981D:B01	ES 164
M00040003A:G10	ES 164
M00040016C:E07	ES 164
M00040023B:B10	ES 164
M00040025A:B04	ES 164
M00040034A:E06	ES 164
M00040034B:G02	ES 164
M00040041A:G08	ES 164
M00040041D:F01	ES 164
M00040045B:H07	ES 164
M00040061C:C08	ES 164
M00040075B:A05	ES 164
M00040078A:C07	ES 164
M00040079B:F06	ES 164
M00040079D:D09	ES 164
M00040081C:E02	ES 164
M00040094B:C08	ES 164
M00040118D:C05	ES 164
M00040123C:A10	ES 164
M00040127C:D02	ES 164
M00022698C:D10	ES 164
M00022702D:E02	ES 164
M00022703D:B11	ES 164
M00022706D:G08	ES 164
M00022727A:G01	ES 164
M00022738D:G08	ES 164
M00022740C:H11	ES 165
M00022797D:A06	ES 165
M00022801D:D09	ES 165
M00022805B:A10	ES 165
M00022812A:G01	ES 165
M00022820A:F07	ES 165
M00022835C:A09	ES 165
M00022854C:G07	ES 165
M00022856D:A07	ES 165
M00022857B:A09	ES 165
M00022897B:F06	ES 165
M00022901A:C05	ES 165
M00022904C:D04	ES 165
M00022924B:A05	ES 165

Clone Name	Tube
M00022924C:F04	ES 165
M00022945A:H09	ES 165
M00022945B:F11	ES 165
M00022947B:D02	ES 165
M00022952A:B02	ES 165
M00022953B:D06	ES 165
M00022964A:B03	ES 165
M00022972C:E05	ES 165
M00022992A:H06	ES 165
M00022992B:G12	ES 165
M00022995C:G07	ES 165
M00023004C:A01	ES 165
M00023007D:D03	ES 165
M00023020C:H03	ES 165
M00023097D:B08	ES 165
M00039184D:H09	ES 165
M00039364D:E05	ES 165
M00039377B:E05	ES 165
M00039377B:H09	ES 165
M00039483A:D10	ES 165
M00039526A:A08	ES 165
M00039537A:F08	ES 165
M00039564D:D04	ES 165
M00039594C:B06	ES 165
M00039598A:E04	ES 165
M00039630D:B07	ES 165
M00039642A:A08	ES 165
M00039642C:F08	ES 165
M00039646A:E06	ES 165
M00039647A:A02	ES 165
M00039647B:A02	ES 165
M00039739B:H12	ES 165
M00040132A:H09	ES 165
M00040162A:E02	ES 165
M00040169A:G06	ES 165
M00040173D:A04	ES 165
M00040174D:G06	ES 165
M00040198A:F12	ES 165
M00040224C:F06	ES 165
M00040247D:D02	ES 165
M00040252C:G05	ES 165
M00040267D:A12	ES 165
M00040287A:C11	ES 165
M00040287C:F10	ES 165
M00040289D:C06	ES 165

Clone Name	Tube
M00039747B:B06	ES 165
M00039748C:G09	ES 165
M00040201A:H01	ES 165
M00040219B:B07	ES 165
M00040291A:G10	ES 165
M00040298B:B09	ES 165
M00040314B:D07	ES 165
M00040326B:G09	ES 165
M00040329A:H05	ES 165
M00040338A:B10	ES 165
M00040344C:D05	ES 165
M00040349D:D07	ES 165
M00040351A:C08	ES 165
M00040351D:G07	ES 165
M00040366B:H10	ES 165
M00040367A:C08	ES 165
M00040381A:B06	ES 165
M00040384B:E04	ES 165
M00040391A:G05	ES 165
M00042525B:H01	ES 165
M00042528C:H01	ES 165
M00042554A:D01	ES 165
M00042557D:B06	ES 165
M00042560C:G06	ES 165
M00042579A:D09	ES 165
M00042719A:G08	ES 165
M00042722C:C09	ES 165
M00042724A:G06	ES 165
M00042732B:H06	ES 165
M00042734A:F05	ES 165
M00042742B:E04	ES 165
M00042743D:G10	ES 165
M00042891C:G08	ES 165
M00042894C:A11	ES 165
M00042908A:F09	ES 165
M00042915B:G11	ES 165
M00054793B:A06	ES 165
M00054911D:E06	ES 166
M00055430A:A01	ES 166
M00055433D:G03	ES 166
M00055448B:E05	ES 166
M00055454A:D02	ES 166
M00055456C:H06	ES 166
M00055466A:F06	ES 166
M00055468A:A08	ES 166

Clone Name	Tube
M00055527B:E01	ES 166
M00055639A:E06	ES 166
M00055653C:B07	ES 166
M00055676A:G02	ES 166
M00055724B:E04	ES 166
M00055724D:C07	ES 166
M00055725D:D09	ES 166
M00055735A:H08	ES 166
M00055745B:A08	ES 166
M00055757A:B01	ES 166
M00055794A:E10	ES 166
M00055805A:H02	ES 166
M00055809A:B09	ES 166
M00055810C:D03	ES 166
M00055818B:D01	ES 166
M00055873D:C02	ES 166
M00055880B:H10	ES 166
M00055919B:C10	ES 166
M00055925D:B07	ES 166
M00055961C:B10	ES 166
M00055975B:F09	ES 166
M00055980C:B04	ES 166
M00056004B:C05	ES 166
M00056024B:F09	ES 166
M00056035D:A08	ES 166
M00056057C:F06	ES 166
M00056105A:D06	ES 166
M00056133A:E11	ES 166
M00056215D:F02	ES 166
M00056217D:E10	ES 166
M00056220D:G02	ES 166
M00056230D:E07	ES 166
M00056244A:B06	ES 166
M00056244C:H05	ES 166
M00056304A:H05	ES 166
M00056320B:A03	ES 166
M00056342A:C03	ES 166
M00056345D:A04	ES 166
M00056436C:F01	ES 166
M00056458C:E01	ES 166
M00042350A:A05	ES 166
M00042433A:E11	ES 166
M00042462B:C02	ES 166
M00042512D:D10	ES 166
M00042766C:D05	ES 166

Clone Name	Tube
M00042788A:F04	ES 166
M00042794A:F01	ES 166
M00042796A:A10	ES 166
M00042801C:D01	ES 166
M00042822A:H04	ES 166
M00042857C:E01	ES 166
M00042858C:G11	ES 166
M00042860B:C07	ES 166
M00042863D:F09	ES 166
M00042878D:F05	ES 166
M00042878D:G06	ES 166
M00042352B:A04	ES 166
M00042352D:B03	ES 166
M00042449B:F05	ES 166
M00042457C:B06	ES 166
M00042516B:D01	ES 166
M00042520B:H04	ES 166
M00043299A:B10	ES 166
M00043306D:C01	ES 166
M00043313D:E09	ES 166
M00043328C:E04	ES 166
M00043336D:B03	ES 166
M00043339C:F11	ES 166
M00043355A:D07	ES 166
M00043358C:A02	ES 166
M00043402B:G07	ES 166
M00054499A:C08	ES 166
M00054528B:E05	ES 166
M00054536B:B01	ES 166
M00054538D:C12	ES 166
M00054542B:A10	ES 166
M00054548C:H06	ES 166
M00054569A:B07	ES 166
M00054579A:C02	ES 166
M00054599D:B03	ES 166
M00054623C:F05	ES 166
M00054643D:F07	ES 166
M00054675D:G03	ES 166
M00054682B:H02	ES 166
M00054683D:G11	ES 166
M00054686A:A09	ES 166
M00054686A:F10	ES 166
M00054693A:E11	ES 166
M00054708C:B06	ES 167
M00054714B:G10	ES 167

Clone Name	Tube
M00054725C:D09	ES 167
M00054744C:F12	ES 167
M00054781B:H04	ES 167
M00054781D:A11	ES 167
M00054786C:D08	ES 167
M00054807D:C11	ES 167
M00054817D:A11	ES 167
M00054818B:F10	ES 167
M00054843A:C01	ES 167
M00054856C:D03	ES 167
M00054866B:C08	ES 167
M00054890C:D05	ES 167
M00054908C:A01	ES 167
M00054931D:E10	ES 167
M00054973B:E12	ES 167
M00054978C:F01	ES 167
M00055001C:G10	ES 167
M00055002B:E08	ES 167
M00055004C:H05	ES 167
M00055023A:E11	ES 167
M00055043B:H08	ES 167
M00055055C:F01	ES 167
M00055081A:A05	ES 167
M00055093B:A03	ES 167
M00055108B:A02	ES 167
M00055117A:E02	ES 167
M00055166C:D10	ES 167
M00055221C:H11	ES 167
M00055232A:E08	ES 167
M00055239D:F11	ES 167
M00055240A:A08	ES 167
M00055244B:F07	ES 167
M00055254A:H03	ES 167
M00055337B:C04	ES 167
M00055375C:F12	ES 167
M00055387C:C12	ES 167
M00055391B:C07	ES 167
M00055395D:D11	ES 167
M00055402A:H01	ES 167
M00055420A:E06	ES 167
M00055423A:B08	ES 167
M00055423C:G12	ES 167
M00055423C:H10	ES 167
M00055424B:H06	ES 167
M00055424D:G05	ES 167

Clone Name	Tube
M00055425C:A04	ES 167
M00055473C:F02	ES 167
M00055477D:B01	ES 167
M00042585A:H11	ES 167
M00042585D:D03	ES 167
M00042585D:E10	ES 167
M00042586A:B01	ES 167
M00042588C:E02	ES 167
M00042621C:C04	ES 167
M00042951D:G12	ES 167
M00042960B:C06	ES 167
M00042967D:C01	ES 167
M00042970C:B01	ES 167
M00042972C:F04	ES 167
M00042976D:C01	ES 167
M00042982D:A10	ES 167
M00042986D:E03	ES 167
M00042996B:H08	ES 167
M00043013B:E03	ES 167
M00043015D:D05	ES 167
M00043016B:F09	ES 167
M00043017C:D08	ES 167
M00043063C:H05	ES 167
M00043070A:C03	ES 167
M00043113C:G09	ES 167

Clone Name	Tube
M00042617B:E01	ES 167
M00043074C:D07	ES 167
M00043076D:A02	ES 167
M00043077B:F11	ES 167
M00043077C:D12	ES 167
M00043077C:G10	ES 167
M00043099A:H04	ES 167
M00043101D:G11	ES 167
M00043134A:F05	ES 167
M00043152C:B10	ES 167
M00043213A:D05	ES 167
M00043219C:C02	ES 167
M00043221D:C12	ES 167
M00043222C:B06	ES 167
M00043455B:C08	ES 167
M00043465C:H11	ES 167
M00043470A:C10	ES 167
M00043485C:C03	ES 167
M00043490C:F02	ES 167
M00043495C:H05	ES 167
M00043528A:E11	ES 167
M00043529A:B08	ES 167
M00043640A:B01	ES 167

CLAIMS

We claim:

1. A library of polynucleotides, the library comprising the sequence information of at least one of SEQ ID NO:1-3351.

2. The library of claim 1, wherein the library is provided on a nucleic acid array.

3. The library of claim 1, wherein the library is provided in a computer-readable format.

4. The library of claim 1, wherein the library comprises a polynucleotide corresponding to a gene differentially expressed in a cancer cell of high metastatic potential relative to a control cell, wherein the control cell is a normal cell or a cell of low metastatic potential, wherein the expression is greater in the metastatic tissue, and wherein the sequence is selected from the group consisting of SEQ ID NOs:14, 137, 151, 152, 171, 200, 254, 262, 271, 348, 412, 472, 507, 520, 530, 588, 623, 637, 660, 678, 680, 700, 714, 774, 812, 834, 901, 937, 976, 1168, 1333, 1352, 1520, 1524, 1546, 1550, 1574, 1580, 1590, 1599, 1607, 1622, 1706, 1752, 1768, 1769, 1780, 1781, 1799, 1803, 1811, 1851, 1856, 1867, 1872, 1875, 1884, 1919, 1923, 1939, 1975, 2024, 2045, 2060, 2071, 2118, 2119, 2128, 2135, 2177, 2181, 2184, 2185, 2190, 2193, 2232, 2239, 2283, 2311, 2314, 2338, 2378, 2393, 2394, 2395, 2398, 2460, 2490, 2505, 2514, 2540, 2542, 2597, 2607, 2640, 2657, 2669, 2670, 2674, 2679, 2684, 2707, 2724, 2757, 2776, 2804, 2818, 2906, 2959, 2964, 2968, 2976, 2980, 2987, 3010, 3043, 3047, 3050, 3071, 3072, 3092, 3095, 3097, 3140, 3157, 3173, 3187, 3203, 3210, 3212, 3220, 3236, 3249, 3264, 3284, 3288, 3305, 3309, 3318, 3330, 3331, and 3335.

5. The library of claim 1, wherein the library comprises a polynucleotide corresponding to a gene differentially expressed in normal colon tissue relative to colon cancer tissue, wherein the expression is greater in the cancer tissue, and wherein the sequence is selected from the group consisting of SEQ ID NOs:7, 164, 734, 836, 928, 965, 987, 1026, 1044, 1119, 1226, 1227, 1251, 1316, 1429, 1442, 1540, 1553, 1560, 1577, 1588, 1610, 1620, 1626, 1673, 2416, 2749, 2976, 3129 and 3132.

6. The library of claim 1, wherein the library comprises a polynucleotide corresponding to a gene differentially expressed in normal colon tissue relative to colon cancer tissue, wherein the expression is greater in normal tissue than cancer tissue, and wherein the sequence is selected from the group consisting of SEQ ID NOs:105, 198, 465, 489, 745, 859, 976, 1011, 1045, 1138, 1226, 1251, 1253, 1392, 1474, 1559, 1571, 1589, 1591, 1607, 1608, 1643, 1753, 1764, 1766, 1782, 1811, 2749, 2784, 2790, 2805, 2976, 3128, 3129, 3146, 3150, and 3151.

7. The library of claim 1, wherein the library comprises a polynucleotide corresponding to a gene differentially expressed in normal human prostate cells relative to human prostate cancer cells, wherein the expression is greater in normal cells than cancer cells, and wherein the sequence is selected from the group consisting of SEQ ID NOs:53, 446, 1410, 1754, 1801, 1845, 2060, 2143, 2632, 2899, and 3338.

8. The library of claim 1, wherein the library comprises a polynucleotide corresponding to a gene differentially expressed in normal human prostate cells relative to human prostate cancer cells, wherein the expression is greater in cancer cells than normal cells, and wherein the sequence is selected from the group consisting of SEQ ID NOs:86, 93, 687, 1269, 1581, 1647, 1649, 1710, 1717, 1772, 1960, 2987, 3128, 3132, 3150, 3222, and 3268.

9. An isolated polynucleotide comprising a nucleotide sequence having at least 90% sequence identity to an identifying sequence of SEQ ID NOs:1-3351 or a degenerate variant or fragment thereof.

10. A recombinant host cell containing the polynucleotide of claim 9.

11. An isolated polypeptide encoded by the polynucleotide of claim 9.

12. An antibody that specifically binds a polypeptide of claim 11.

13. A vector comprising the polynucleotide of claim 9.

14. A method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, the method comprising the step of:

detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, wherein the gene product is encoded by a

gene corresponding to a sequence of at least one of SEQ ID NOs: 14, 137, 151, 152, 171, 200, 254, 262, 271, 348, 412, 472, 507, 520, 530, 588, 623, 637, 660, 678, 680, 700, 714, 774, 812, 834, 901, 937, 976, 1168, 1333, 1352, 1520, 1524, 1546, 1550, 1574, 1580, 1590, 1599, 1607, 1622, 1706, 1752, 1768, 1769, 1780, 1781, 1799, 1803, 1811, 1851, 1856, 1867, 1872, 1875, 1884, 1919, 1923, 1939, 1975, 2024, 2045, 2060, 2071, 2118, 2119, 2128, 2135, 2177, 2181, 2184, 2185, 2190, 2193, 2232, 2239, 2283, 2311, 2314, 2338, 2378, 2393, 2394, 2395, 2398, 2460, 2490, 2505, 2514, 2540, 2542, 2597, 2607, 2640, 2657, 2669, 2670, 2674, 2679, 2684, 2707, 2724, 2757, 2776, 2804, 2818, 2906, 2959, 2964, 2968, 2976, 2980, 2987, 3010, 3043, 3047, 3050, 3071, 3072, 3092, 3095, 3097, 3140, 3157, 3173, 3187, 3203, 3210, 3212, 3220, 3236, 3249, 3264, 3284, 3288, 3305, 3309, 3318, 3330, 3331, and 3335.

wherein detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived.

15. A method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, the method comprising the step of:

detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, wherein the gene product is encoded by a gene corresponding to a sequence of at least one of SEQ ID NOs:7, 164, 734, 836, 928, 965, 987, 1026, 1044, 1119, 1226, 1227, 1251, 1316, 1429, 1442, 1540, 1553, 1560, 1577, 1588, 1610, 1620, 1626, 1673, 1960, 2416, 2749, 2976, 2987, 3128, 3129, 3132, 3150, 3222, and 3268.

wherein detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived.

SEQUENCE LISTING

<210> 1 <211> 415 <212> DNA <213> Homo sapien
ttcgaattcg gcacgagatt tcatagatgg agaaactgat cacagagctg taatgaagac 60
agaattgaga tatgagggca aaagctaatt aaacgcatcc tcacaggtag ccttttcttc 120
agtgaacctg tagactagtc cagtaatact tattaaaatt agttgttaga ggctgggcat 180
ggtggttcaa gcctgtaatc tcagcactgt gggaggccaa ggcggacaga tcactcagag 240
tcagaagttc gagaccagct tggccaacat ggcaaaaccc tgtctctact aaaaatacaa 300
aaattagttg ggtgtggtgg cacatgcctg taatcccagc cactcgggag gtgaaggcac 360
aagaattggt tgaacctggg aagcagaggt tgcagtgagc tgagattgca ctgct 415
<210> 2 <211> 225 <212> DNA <213> Homo sapien
ggcacgagct ctctctctct ctcncncnaa ctctctgtct ctctctctct ctaggctctc 60
tctctctctc tctctatcta tctctcagac tatgtgtgag tgtgagagag agagagagag 120
agagagagag agagagagag agaaagacag agagagacag gatgaatagt ataaaagagg 180
gggggctaga gaaagagaga aggaaaaaag agagaaaaaa aaagc 225
<210> 3 <211> 437 <212> DNA <213> Homo sapien
ggcacgagag agactgtggc tcatgcttgt gatccccctg ccttggcctc ccaaagttct 60
gggattacag gggtgaacca ctgtgcctgg cccatttttc tttataaata ttgcaacata 120
atgttttata gacaaacatt caagggtact ttggctttat gaacttcagg atttctggtg 180
ctagaaaagc gcttgaagca gtatcaccaa gattttagat attaaaaagt ctggtgtacc 240
agacattgag tcataatcat ctatattcaa gggatacttt cattgataac tttggtatta 300
tgctgccctt cacagaagac aacgtctcgg gcaggatcac atgctcccta gcagatgctg 360
atcagtgatg tcatagaaat tacatgaatg catttgcttt aaatagcagt taaccattgt 420
atatggggcg ttttgct 437
<210> 4 <211> 360 <212> DNA <213> Homo sapien
ggcacgaggc ctggcatggt ggcacatgcc cataattcca gctactcggg aggctaaggc 60
aggagaatcg cttgaacctg acggggtgga ggttgcagtg agccgagatc gcaccacttc 120
actccagcct gggcgaaaga gcgaaactcc atctcaaaaa aaaaaaaggg aaggggaaaa 180
aaaaccggaa aagatttggt tggggaactt ttaggagggg tggggccctt ggggccctta 240
actaacccca gggaatcctt taaagggaaa gggggggaag ggttgtcaaa ccccgggggg 300
tcatggtaaa aaaagggttg ggttccctta attctttccc caattttcaa aacccataaa 360
<210> 5 <211> 600 <212> DNA <213> Homo sapien
tacggctgcg agaagacgac agaagggtgg ctaacacggt gaaaccccgt ctgtactaaa 60
aatacaaaaa gttagccggg cgtggtggcg ggcgcctgta gtctcagcta cttgggaggc 120
tgaggcagag gcaggagaat ggtgtgaacc tgggagacgg aggttgtggt gagccgagat 180
caggccactg cactccagcc tgggtaacag agcaagactc cgtctcaaaa aaaaaaaaaa 240
aaaaaaaaaa aggggggggg gttttttttcc gtaacccccа ccttgaaaaa acccttтggg 300
ggttggggcc ccccccccct taaggggggg gaaaaaaagg ttttttttgg gaaaattggg 360
gggctttttt ttttttgga ccccttttaa ggcggaaaaa cctgttaacc acaaatttgg 420
ttttttttt ttttgtttg ggggggggg ggaggggttt tnnnnnnnnn ncnangaaag 480
gggggcccc aacacggtgt ggttttaatc ccccttaggg cggccccttt ttttttttgg 540
gggcgcgcgg tgggggggaa gaaaaaatgg ggnttttgtg ttaccctgta ctattttaac 600
<210> 6 <211> 404 <212> DNA <213> Homo sapien
attcggcacg aggagagaga gagagagaga gagagagaga gagagagaga gagagagaga 60
gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga 120
gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga 180
gagagagaga gagagttttt ttttttttt taaaaaaata ttttttttt tgcgcgcaca 240
cacactctct ctttttttt ttttttttt acactccgcg cgcccgcttt atatacaccc 300
acacatatat atatatatat atatatatat atgtgtatat atcttttttt taccccacc 360
cgccgggggc gcgcgcacgc cctccccccc ctctgtctct attt 404
<210> 7 <211> 358 <212> DNA <213> Homo sapien
tacggctgcg agaagacgac agaaggggct ggtaattttt gtatttttag tagagactgg 60
gtttcaccat gttggccagg ctggtcttga actccaggcc tcgagtaatc cacccacctt 120
ggcctcccaa agtgttgcga ttagaggcat gagccaccgt gctcaggctt cccacaataa 180

attcggcacg agtggaggcc ccggagaccc caggagagcc accactttct cctgggttct 60
gaacacagcc caggtgggaa caatgctgcc cctcatgatg aagtggcctg tgtggcttga 120
gcgccccata gtccccagtc agagcagagt ggtgtcccca gatgacttca gaccccatag 180
ctgggcaaga tgcgcttgtt ttggactctg cgctgagcag aaccagctcc cccaactcct 240
gcagatagag aactgacttc cgagagctgt aggtgaagtg aggaccaggc agcagtccag 300
agctgtgagg ccccaggccc agaggaatgg aatgaagaaa gacctgttcc acacaaggag 360
gggttttcta gtggaagctg agcttggaag ctcctg 396

<210> 171 <211> 390 <212> DNA <213> Homo sapien

ggcacgagga gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga 60
gagagagaga gagagagaga gagagagaga gagagagaga gagcgcgcct ctggcacact 120
ctctctctct acacactctg tctgtgcgcg ctccacactc tatataccgc acacacgctc 180
agagtgtctc cgcgcgcgcg cgcgccaaga cactctagtg cgcgcgtatt tgtgcgctct 240
ctctctcccc ccccacgcgc gcgccacaaa actctctttt tggcgctctc tggcacacac 300
actctcttct ctatgcgcac tctctctctg agtctctctc tcttatatat acccgcgcga 360
tacatatctg tgtgcgagac tctgtgtgcg 390

<210> 172 <211> 399 <212> DNA <213> Homo sapien

ggcacgagct accctccacg ggagacgaag aggtgtttgt ttccggctcc accccacctc 60
ccagctgtgc cgtgcggagc tgcctctctg ccagtgccct ccaggctctg acccagtctc 120
cgctgctgtt ccaggggaaa acaccttcct ctcagagcaa agaccccaga gatgaggatg 180
tggatgttct tccctccact gtagaagact ctcctttcag tcgcgctttc tccaggaggc 240
gccccatcag cagaacttat acacggaaga agctcatggg aacctggctg gaggacttat 300
agccacaaac attactgagc ccaaaagatc aaggagtcag ccaggaccct gtggacataa 360
agaagttgga tgcctggtcc caagcctctt ttgccatgg 399

<210> 173 <211> 396 <212> DNA <213> Homo sapien

gaattcggca cgagcccagt ggtgccaggg cagagtcccc ctccctgacc tgacttgtgc 60
acctcgtcac ccaccgccag cagtgtcccc ccacaacagg cttgctcagt acagcaccca 120
acccaagtcc ccagcaccca caccccagtg agtttcctgt gccctatagg ctcagctgct 180
tctcgtcctc cccccacttg ggatccttgg aacagggagt ggttcttatt taggtccctg 240
aggtaccaag cacaggcttt gctcttagca gccgccactc cagtgatgaa gccgttagca 300
gactggcctc tgcagagctc tgcggggagg tgcctggctt ctccggcctc caccctggcc 360
cagagctgcc tcctgagcag cggatcccaa cctgcg 396

<210> 174 <211> 383 <212> DNA <213> Homo sapien

ggcacgagcc caggtctctc atgagaaact tgtttaccct cttagatacc cttgagtctc 60
ttgtctgtgt ctggtgtatt tatttattta gcctaccaag atagccactc ttcaggagag 120
ttctgaattt ggaaagaagt taggatcagg tgtgttggtc aagtgagaca cagaggaggc 180
cactcaacaa aacccatgaa ataccagaag cagtgagttc ctcgcaggtc cagagagaag 240
agggcagcac gctggactgg gggagccgtc aggacccttg tgctcgccag caggtgggga 300
gcaagagaga tggagtgtgg gccctgagag ctgaagcctt tatggggtcc aggccatcac 360
cccagcaggt tcccaagaag ttg 383

<210> 175 <211> 386 <212> DNA <213> Homo sapien

ggcacgaggg caagagattc tccactgcta tgggcctcac aagagccgga tgggggttgc 60
cgaaaggcag cagaagctga ggtctcagta tttctttgac tgcgcctgtc cagcttgtca 120
aactgaggca cacaggatgg ctgcagggcc caggtgggaa gcattctgtt gcaacagttg 180
cggagcgccc atgcagggag atgacgtgct gcgctgtggc agcagatctt gtgcagaatc 240
cgccgtcagc agggaccacc tggtctctcg gttacaggac ctacagcagc aggtcagagt 300
ggcccagaag cttctcagag atggtgaact aaagcgagct gttcagcggc tgtcggggtg 360
ccagcgtgac gccgagagct tcctgt 386

<210> 176 <211> 383 <212> DNA <213> Homo sapien

catcgattcg aattcggcac gagtgacaat gttgtcctcc tgttcatctg tgcaccactt 60
gacagactgt agcttctctt gctctcgacc ggccctgcat tcttccgcac cctccctagc 120
tctgaaatca actctcttcg gtcgtatcca ccttgcaccc gcaagtcaag ccgccccttg 180
tagaaaaatc cctccacctt ccgttccccg ctaggtcaac cccactgtag acaggaaagc 240
caggccagga gagtccgaat gagaatttat tgtgaatcga ttcccaagct cccttccggg 300
acaagtggtc tgggacaggg aggagcaacg gccccagcgc gcaacgctct gcgcgttcct 360
cccgaatccc gtcgcttctc gac 383

<210> 177 <211> 393 <212> DNA <213> Homo sapien

agacctcttt tgggaaggct gcaagggagg gtcacaacat gcatctaaag tgtaaaaatt 360
aaagttttcc tttcaaaata catttgactt cctcttcatg taaggg 406
<210> 775 <211> 402 <212> DNA <213> Homo sapien
ggcacgagga gagagagaga gagagagtgt tgtagtgaga gagagagaga gagagagaga 60
gagagagaga gagagagaga gagagagaga gagagagagt gagagagaga gagagacaga 120
gagagagaga gagagagaga gagagagaga gagagagaga gagagagaga gagtgttttt 180
ttttttttctc tcacacaccc ttttttctct ctctgtgtgt gttttttttt gtcagactct 240
ttttttcttcc ctccccccgcc cgcgagattc ttttttttag cactctctct ctcttccctc 300
tttttgtgtc ccacatattt tttctcgcgc gcttcccccc ccttgtgcgt gtgttttttt 360
ctctcacgcg cgcgtgtttt ttattttgtc tctctctccc cg 402
<210> 776 <211> 407 <212> DNA <213> Homo sapien
tcgattcgaa ttcggcacga gaagaactag aggagaaaat gtcacaagca agacaaatct 60
gcccagagcg tatagaagta gaaaaatctg catcaattct ggacaaagaa attaatcgat 120
taaggcagaa gatacaggca gaacatgcta gtcatggaga tcgagaggaa ataatgaggc 180
agtaccaaga agcaagagag acctatcttg atctggatag taaagtgagg actttaaaaa 240
agtttattaa attactggga gaaatcatgg agcacagatt caagacatat caacaattta 300
gaaggtgttt gactttacga tgcaaattat actttgacaa cttactatct cagcgggcct 360
attgtggaaa aatgaatttt gaccacaaga atgaaactct aagtata 407
<210> 777 <211> 405 <212> DNA <213> Homo sapien
attcggcacg agaagaacta gaggagaaaa tgtcacaagc aagacaaatc tgcccagagc 60
gtatagaagt agaaaaatct gcatcaattc tggacaaaga aattaatcga ttaaggcaga 120
agatacaggc agaacatgct agtcatggag atcgagagga aataatgagg cagtaccaag 180
aagcaagaga gacctatctt gatctggata gtaaagtgag gactttaaaa aagtttatta 240
aattactggg agaaatcatg gagcacagat tcaagacata tcaacaattt agaaggtgtt 300
tgactttacg atgcaaatta tactttgaca acttactatc tcagcgggcc tattgtggaa 360
aaatgaattt tgaccacaag aatgaaactc taagtatatc atatg 405
<210> 778 <211> 393 <212> DNA <213> Homo sapien
ggcaccagag ccaccacacc tggctaggtt tacatttttа gaatatccct tggaaagtgg 60
ttggagagta gcaaaagtgt gttgtttggt aaaatatctc tggaaggaaa cttcagacaa 120
tagtaacagc agtcttcttg gcaggcaacc tgggagacag ggataaatgg gagactccct 180
gtttataaca tacccctttg tactttctaa gttttatact atgtacatgt attcattgac 240
tgaataaata gctttataaa gtcgtttttа taaaagagaa ggttgggagg agctatcagg 300
tagcaactgc agatgtctaa ggaagaggtc atggtggtca tttggactgg gtgctggtgg 360
tggagtcaaa gtggaccaag tcaagagact ttt 393
<210> 779 <211> 387 <212> DNA <213> Homo sapien
agatttcttt caattggtct tcccattgca gttactgtta tttctctttt ttggttaact 60
ttaaatcaaa actcaaaata tgttcatcca gagtgtgtct taagtaactt acgtgtctta 120
agtaacaggg accagagaca tgttacctac aagagttctg ggctatcctt ttcattctta 180
tcacatatca tagcttgaat attacaacag tgtgggagag aatcaaccgt aaaaatgtct 240
tcattaatta gacccagtta ttccactttt ggtaatgtct ctcacattga cacagtataa 300
aaattatatg caccaagatg tccaagtgac atacttttag agccaattat anacacttta 360
aagttgggga aagattgcaa ctntttt 387
<210> 780 <211> 386 <212> DNA <213> Homo sapien
ggcacgagcc atcccttata gaagaggtca ttcctgctct tccttctcca tggctagagg 60
atctacatga actatttaga ttttttctac ctgggagatt taactcctct ctcctattta 120
tttatttata tatcagcatg gacttgcagg ccaacagaga ttttgagaaa cacattgaag 180
gatctgttaa cacttgatat acccaataaa agcagtggtt gtgccagtgc tgatctgtct 240
tgatgtgaat gtgaacaatg ggaacctgag ctgagcagtt aaatgtaggg tgacagaaac 300
tggacctctt ccaaaacatg tgacagagta ataccagagc caacttcttc gccaaattaa 360
agtttacaag aattaacctg tcatcn 386
<210> 781 <211> 392 <212> DNA <213> Homo sapien
attcggcacg aggaaaatca gaagccctat tgtatctggt atttcacaac cagacgtttt 60
caatcactac ccttttgctg agtgccatga aactgatagt gatgaatggg tccctcctac 120
cacacaaaaa atatttcctt cagatatgct tggattccaa ggcataggtc tagggaaatg 180
ccttgctgcc tatcatttcc ctgatcaaca agagttacca agaaagaaac tgaaacatat 240
tagacaagga accaataaag gtttaattaa gaagaaatta aagaatatgc ttgcagcagt 300

tcctgtattt ttgggaagcg ccttgaagaa caaaggagtt cagcctcttt tagatgctgt 360
tttagaatac cttccaag 378
<210> 1818 <211> 408 <212> DNA <213> Homo sapien
atcgattcgc tcatctcaga gactggtgga agccatgaca agcgctttgt aatggaggta 60
gaagtagatg gacagaaatt cagaggcgca ggtccaaata agaaagtggc aaaggcgagt 120
gcagctttag ctgccttgga gaaactgttt tctggaccca atgcggcaaa taataagaaa 180
aagaagatta tccctcaggc aaagggcgtt gtgaatacag ctgtgtctgc agcagtccaa 240
gctgttcggg gcagaggaag aggaactcta acaaggggag cttttgttgg ggcgacagct 300
gctcctggct acatagctcc aggctatgga acaccatatg gttacagcac agctgccccc 360
gcctatggtt tacccaagag aatggttctg ttacccgtta tgaaattt 408
<210> 1819 <211> 386 <212> DNA <213> Homo sapien
tacggctgcg agaagacgac agaaggggaa aatttagaag accttgaaat aatcattcaa 60
ctgaagaaaa ggaaaaaata caggaaaact aaagttccag ttgtaaagga accagaacct 120
gaaatcatta cggaacctgt ggatgtgcct acgtttctga aggctgctct ggagaataaa 180
ctgccagtag tagaaaaatt cttgtcagac aagaacaatc cagatgtttg tgatgagtat 240
aaacggacag ctcttcatag agcatgcttg gaaggacatt tggcaattgt ggagaagtta 300
atggaagctg gagcccagat cgaattccgt gatatgcttg aatccacagc catccactgg 360
gcaagccgtg gaggaaacct tgatgt 386
<210> 1820 <211> 402 <212> DNA <213> Homo sapien
ggcacgagag gacaaagaga ggccggatca aaccaacccc tccgccaact ggctgcacgc 60
tcgctcttcc cggaaaaagc gctgtcccta caccaaatac cagacgctgg agctagagaa 120
ggagtttctc ttcaatatgt acctcaccaa ggaccgtagg cacgaagtgg ccagactcct 180
caatctgagt gagagacaag tcaaaatctg gtttcagaac cggcggatga aaatgaagaa 240
aatgaataag gagcagggca aagagtaaag attaaagatt acccccagtc ctccctagct 300
cttccccatc tcactcttag ttatgtgacg actgcaaagc cagtgctgtc tgggatgtat 360
tcaagtgaat ggggaaggga gtctctcttc caagtccttt an 402
<210> 1821 <211> 398 <212> DNA <213> Homo sapien
ggcacgagag gacaaagaga ggccggatca aaccaacccc tccgccaact ggctgcacgc 60
tcgctcttcc cggaaaaagc gctgtcccta caccaaatac cagacgctgg agctagagaa 120
ggagtttctc ttcaatatgt acctcaccag ggaccgtagg cacgaagtgg ccagactcct 180
caatctgagt gagagacaag tcaaaatctg gtttcagaac cggcggatga aaatgaagaa 240
aatgaataag gagcagggca aagagtaaag attaaagatt acccccagtc ctccctagct 300
cttccccatc tcactcttag ttatgtgacg actgcaaagc cagtgctgtc tgggatgtat 360
tcaagtgaat ggggaaggga gtctctcttc caagtccn 398
<210> 1822 <211> 367 <212> DNA <213> Homo sapien
cgttgctgtc ggtccagaaa gtagaatgct gtgcatcgct ggagtttcag ctcatgtcat 60
tatttataga ttcagcaagc aggaagtaat cacagaagtc attccgatgc ttgaagttcg 120
attattatat gagataaatg atgtggaaac tccggagggt gagcagccac cacctttgcc 180
aacacccgtg ggagggtcca accctcagcc catccctcct cagtctcatc catctaccag 240
tagcagttca tctgatgggc ttcgtgataa tgtaccttgt ttaaaagtta aaaactcacc 300
acttaaacag tctccaggtt atcaaacaga actagttatt cagttggttt gggtgggtgg 360
agaacca 367
<210> 1823 <211> 370 <212> DNA <213> Homo sapien
tacggctgcg agaagataca naagnagacc ttcttcgtgc tcagggcctg ggagatatta 60
ttgatacatc catggggtcc ctcacttcat ccccatcttc ctgctcactc agtagtcagg 120
tgggcttgac gtctgtgacc agtattcaag agaggatcat gtctacacct ggaggagagg 180
aagctattga acgtttaaag gaatcagaga agatcattgc tgagttgaat gaaacttggg 240
aagagaagct tcgtaaaaca gaggccatca gaatggagag agaggctttg ttggctgaga 300
tgggagttgc cattcgggaa gatggaggaa ccctaggggt tttctcacct aaaaagaccc 360
cacatcttgt 370
<210> 1824 <211> 447 <212> DNA <213> Homo sapien
tacggctgcg agaagacgac agaaggggtt attttgcaag cgggaggggc cgtgcgcgct 60
cctgcctcag gcctctgtcc cccacccccc ttccccggtc ccaggctctc cttcggaaag 120
atgtcggaca cggcagtagc tgatacccgg cgccttaact cgaagccgca ggacctgacc 180
gacgcttacg ggccgccaag taacttcctg gagatcgaca tctttaatcc tcaaacggtg 240
ggcgtgggac gcgcgcgctt caccacctat gaggttcgca tgcggacaaa cctacctatc 300